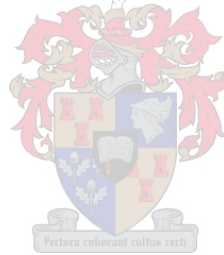


The effect of rater-ratee personality similarity on ratings of  
task-oriented work behaviours

By

Antonette Botes

*Thesis presented in fulfillment of the requirements for the degree Master of Commerce in  
Industrial Psychology at the University of Stellenbosch*



Supervisor: François de Kock  
Faculty of Economic and Management Sciences  
Department of Industrial Psychology

December 2012

## DECLARATION

I herewith declare this work to be my own, that I have acknowledged all the sources I have consulted in the document itself and not only in the bibliography, that all wording unaccompanied by a reference is my own, and that no part of this assignment/essay has been directly sourced from the internet without providing the necessary recognition.

I acknowledge that if any part of this declaration is found to be false I shall receive no marks for this assignment / essay, shall not be allowed to complete this module, and that charges can be laid against me for plagiarism before the Central Disciplinary Committee of the University.

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Copyright © 2012 Stellenbosch University

All rights reserved

## ABSTRACT

As a means to measure job performance, performance appraisal plays a central role in effective individual and organisational management (Behn, 2003). Sound performance management and performance measurement are fundamental to a productive workplace and critical for a high-performing organisation (Jordan, 2002). Performance appraisal research has shifted its emphasis from psychometric issues to the examination of rater cognitive processes and the social and contextual variables which affect performance evaluation. Since raters are important factors in successful performance measurement, one line of research has investigated the effect of similarity, between rater and ratee, on subsequent performance ratings. These studies have mostly relied on similarity measures based on physical similarity characteristics, such as demographic variables. The inconclusive nature of these studies' findings suggests that the complexity of interpersonal similarity and its effect on ratings has most likely been oversimplified. In the social-cognition literature, substantial evidence exists that rater-ratee acquaintance shifts the focus of similarity judgment to "deeper", sometimes unobservable, characteristics, like values, motives and attitudes. This research study investigates whether rater-ratee similarity in Big Five personality traits unduly influences task-orientated performance ratings. Self-report personality data (IPIP; Goldberg, 2006) were collected from university lecturers and their students ( $N = 152$ ). Actual lecturer task performance assessment data (end-of-semester student feedback ratings) were gathered concurrently. Data were analysed through polynomial regression analysis and response surface methodology. Results indicated that ratee (i.e., lecturer) extraversion ( $r = .357$ ), conscientiousness ( $r = .413$ ) and openness ( $r = .178$ ) had significant main effects on average performance ratings. Also, rater-ratee personality similarity in extraversion ( $p < .001$ ), neuroticism ( $p < .01$ ) and openness ( $p < .001$ ) had a significant effect on performance ratings, with the effects of agreeableness and conscientiousness also approaching significance. The

present study further extends earlier research by using task performance ratings as criterion measures — as opposed to earlier studies that used contextual performance ratings — and also used “upward” ratings of seniors, instead of peer- or ‘downward’ ratings of performance, as was done in earlier studies of personality similarity effects. The results suggest that (a) earlier conclusions that personality similarity does not affect performance ratings seem to be premature, (b) more research is needed to investigate why personality similarity affects ratings and last, (c) we do not yet understand the boundary conditions that affect this phenomenon.

## ACKNOWLEDGEMENTS

I would like to thank my Lord and Saviour for giving me the opportunity and ability to complete this thesis. You are my solid foundation;

my parents and brothers for all your love, prayers and support;

Regardt, my husband, for inspiration and motivation, for making me want to be the best I can be and for helping me whenever I asked;

my housemates, Karien and Suzaan, for being there through all the ups and downs;

Francois, my study leader, for your guidance, wisdom, and believing in me when I was ready to give up;

the Stellenbosch University Centre for Learning and Teaching (CTL) for helping me with data-capturing – you made my task so much easier;

and my class-mates DB, Christine, Nadia and Marthine. What a journey!

## TABLE OF CONTENTS

DECLARATION.....	i
ABSTRACT.....	ii
ACKNOWLEDGEMENT.....	iv
APPENDICES.....	vii
LIST OF TABLES AND FIGURES.....	viii
<b>1. CHAPTER 1 - INTRODUCTION AND JUSTIFICATION OF STUDY.....</b>	<b>1</b>
<b>2. CHAPTER 2 – LITERATURE REVIEW .....</b>	<b>11</b>
2.1 Introduction.....	11
2.2 Performance Measurement in general.....	11
2.2.1 Performance Appraisal Models .....	13
2.2.1.1 Model proposed by Feldman .....	13
2.2.1.2 Model proposed by DeNisi, Cafferty and Meglino .....	14
2.2.1.3 Summary of performance appraisal models .....	16
2.2.2 Performance appraisal problems .....	17
2.2.3 Generalizability .....	22
2.3 Measurement Error in Performance Measures .....	24
2.3.1 Rater bias in Performance Measurement .....	25
2.3.2 Psychometric research on ratings of performance .....	27
2.4 Theories underlying similarity effects .....	28
2.4.1 Social Identity Theory (SIT) .....	28
2.4.2 The mechanisms of similarity effects .....	29
2.4.3 Similarity as a construct .....	31
2.5 General research findings on similarity effects .....	32
2.5.1 Choice of personality as similarity variable .....	37
2.6 Personality .....	38
2.6.1 Personality as similarity measure .....	39
2.7 Previous studies on personality similarity and their effects on performance rating .....	41
2.8 Summary and conclusion .....	41

2.9	Research question and objective .....	44
2.9.1	Research Question .....	44
2.9.2	Research Objectives .....	45
2.9.3	Present Study .....	45
2.10	Hypotheses .....	47
2.10.1	Personality main effects .....	47
2.10.2	Relational similarity hypotheses.....	47
2.11	An overview of polynomial regression with response surface analysis .....	49
2.12	Conclusion.....	51
<b>3.</b>	<b>CHAPTER 3 – RESEARCH DESIGN AND METHODOLOGY.....</b>	<b>52</b>
3.1	Introduction.....	52
3.2	Statistical hypothesis.....	52
3.3	Research Design .....	53
3.4	Sample .....	54
3.4.1	Participants .....	55
3.5	Measuring Instruments .....	55
3.5.1	Personality measure .....	57
3.5.2	Task-orientated performance rating .....	59
3.5.3	Demographics.....	60
3.6	Procedure .....	61
3.7	Data Analysis .....	61
3.7.1	Step by step overview of polynomial regression with response surface analysis .....	63
3.8	Conclusion .....	65
<b>4.</b>	<b>CHAPTER 4 – RESULTS .....</b>	<b>66</b>
4.1	Introduction.....	66
4.2	Data Cleaning.....	66
4.3	Testing for assumptions .....	67
4.3.1	Normality, linearity, homoscedasticity .....	67
4.4	Assessing the measurement properties .....	72

4.4.1	Subscale level descriptive statistics .....	73
4.4.1.1	Item analysis .....	73
4.4.1.2	Explanatory factor analysis (EFA) .....	73
4.4.2	Basic descriptive statistics of ratee ratings.....	75
4.4.3	Means, standard deviations, and correlation between study variables .....	78
4.5	Testing main effects .....	79
4.6	Frequency of Big 5 personality similarity and dissimilarity in sample (%) .....	81
4.7	Results of polynomial regression analysis .....	82
4.7.1	Extraversion .....	84
4.7.2	Agreeableness .....	86
4.7.3	Conscientiousness .....	88
4.7.4	Neuroticism .....	90
4.7.5	Openness .....	91
4.8	Summary and conclusion .....	93
<b>5.</b>	<b>CHAPTER 5 - DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS ....</b>	<b>95</b>
5.1	Introduction.....	95
5.2	Background of the study .....	95
5.3	Summary of results .....	96
5.4	Discussion of results .....	97
5.5	Limitations of study .....	102
5.6	Practical implications and recommendations for future research .....	103
5.7	Concluding remarks .....	105
	LIST OF REFERENCES.....	106
	APPENDIX A: Mini IPIP Questionnaire.....	116
	APPENDIX B: Lecturer evaluation form.....	117
	APPENDIX C: Consent form for ratee participation in research.....	118
	APPENDIX D: Consent form for rater participation in research.....	121
	APPENDIX E: Ethical permission to conduct research.....	124



## LIST OF FIGURES

Figure 2.1 – Proposed model of the appraisal process .....	15
Figure 2.2 – Path model of classes of variables affecting ratings .....	23
Figure 4.1 – Histogram depicting normality of distribution for the IPIP EXTRAVERSION independent variable .....	68
Figure 4.2 – Histogram depicting normality of distribution for the IPIP AGREEABLENESS independent variable .....	68
Figure 4.3 – Histogram depicting normality of distribution for the IPIP CONSCIENTIOUSNESS independent variable .....	69
Figure 4.4 – Histogram depicting normality of distribution for the IPIP NEUROTICISM independent variable .....	69
Figure 4.5 – Histogram depicting normality of distribution for the IPIP OPPENNESS independent variable .....	70
Figure 4.6 – Graph output showing within-group scatterplot of rater personality dimensions and Average Performance .....	71
Figure 4.7 – Graph output showing within-group scatterplot of ratee personality dimensions and Average Performance .....	71
Figure 4.8 – Graph output showing within-group scatterplot of rater-ratee personality dimensions and Average Performance .....	72
Figure 4.9 – Scree-plot suggesting the amount of factors to extract from performance measure items .....	77
Figure 4.10 – Performance rating as predicted by rater extraversion and ratee extraversion discrepancy .....	86
Figure 4.11 – Performance rating as predicted by rater agreeableness and ratee extraversion discrepancy .....	89
Figure 4.12 – Performance rating as predicted by rater conscientiousness and ratee conscientiousness discrepancy .....	89
Figure 4.13 – Performance rating as predicted by rater neuroticism and ratee neuroticism discrepancy .....	91
Figure 4.14 – Performance rating as predicted by rater openness and ratee openness discrepancy .....	93

**LIST OF TABLES**

Table 2.1 – Summary of findings on interaction of rater-ratee characteristics and performance ratings .....	34
Table 3.1 – Demographic details of sample .....	56
Table 4.1 – Total variance explained .....	76
Table 4.2 – Rotated component matrix of performance rating measure items .....	77
Table 4.3 – Descriptive statistics of each specific ratee’s performance .....	78
Table 4.4 – Means, Standard Deviations, and Correlations of study variables .....	80
Table 4.5 – Frequencies of dyadic similarities between rater and ratee personality dimensions .....	81
Table 4.6 – Results of Polynomial Regression Analysis on Big-5 personality dimensions .....	83
Table 5.1 – Summary of hypotheses and conclusions .....	99

## CHAPTER 1

### 1. INTRODUCTION AND JUSTIFICATION FOR THE STUDY

Performance measurement plays a central role in effective individual and organisational performance management (Behn, 2003). Effectively measuring performance is fundamental to a productive workplace and critical for a high-performing organisation (Jordan, 2002). Firms that seek a competitive advantage through their human resources must be able to measure and manage the behaviour and work results of employees, primarily because sound measurement allows better employment decisions (Noe, Hollenbeck, Gerhart & Wright, 2006).

Performance management can be defined as “the means through which managers ensure that employees’ activities and outputs are congruent with the organisation’s goals” (Noe et al., 2006). A performance management system typically comprises of three parts: defining performance, measuring performance, and feeding back performance information (Noe et al., 2006). During the performance definition phase, aspects of performance that are relevant to the organisation are specified through job analysis. Next, those aspects are measured through performance appraisal and, lastly, feedback is provided to employees.

Performance measurement plays a cardinal role in any successful performance management system. Performance appraisal – the term typically used to describe performance measurement – is defined by Kline and Sulsky (2009) as “a general heading for a variety of activities through which organizations seek to assess employees and develop their competence, enhance performance and distribute rewards”.

There are different approaches to measuring performance (Noe et al., 2005), including the comparative approach (ranking, forced distribution, paired comparison); the attribute approach (graphic rating scale, mixed-standard scale); the behavioural approach (critical incidents, behaviourally anchored rating scales, behavioural observation scales, organisational behaviour modification, assessment centres); the results approach (managing by objectives, productivity measurement and evaluation system); and the quality approach (performance measurement against the fundamental principles, internal peer reviews, external peer reviews). Each of these approaches involves trade-offs between various pros and cons.

Ideally, performance appraisal results should help managers to make formal as well as informal personnel decisions and provide information that will best enable them to improve employee performance (MacDonald & Sulsky, 2009). These formal and informal personnel decisions include, but are not limited to promotions, feedback, development areas and bonus pay. It is, therefore, crucial for these performance assessment results to be as reliable, valid and fair as possible. Unfortunately, the method used to gather the majority of performance assessment data, i.e., subjective ratings, have been shown to be prone to various types of systematic as well as random error (Landy & Farr, 1980), which adversely affect the quality of subjective rating data used for important personnel decisions.

Effective rating remains important to ensure that correct decisions, for the above-mentioned defined needs, are made. However, acknowledging the importance of performance measurement, and actually measuring performance accurately, are two different issues. In what way does one determine whether the correct decisions about performance levels are

made, since most people involved in performance measurement depend on judgemental indices of one type or another (Landy & Farr, 1980)?

Various criteria are often used to assess the effectiveness of performance appraisal outcomes. Murphy and Cleveland (1995) mentioned Jacobs and colleagues' "meaningful framework" for describing the numerous types of appraisal criteria. Their framework contains three categories of important benchmarks, namely, utilisation criteria, qualitative criteria and quantitative measures of effectiveness. First, the *utilisation criteria* address the purpose for which the appraisal is conducted (e.g., administrative decisions, employee development, etc.). Second, *qualitative criteria* take into account the relevance of the appraisal to job performance, data availability, equivalence, interpretability, and practicality. Last, *quantitative measures of effectiveness* are frequently used as appraisal criteria. Since their outline of these criteria, many other yardsticks have been added, including reliability, interrater reliability, and validity.

Another criterion used is the quality of the source of the rating; whether peer, supervisor or subordinate ratings provide the most psychometrically sound and accurate rating, as well as the agreement and equivalence of ratings across these sources (Scullen, Mount & Goff, 2000). Performance appraisal research has evolved through a number of stages over the last few decades. Strauss, Barrick and Connerley (2001) suggest that, in general, research on performance appraisal has evolved from an emphasis on psychometric issues to the examination of the cognitive process of raters and the social and contextual variables which affect performance evaluation. Levy and Williams (2004) also came to a similar conclusion in their systematic review of relevant performance measurement research, including more than 300 articles, which suggests that, as a field, there is much more awareness of the

importance of the *social context* within which the performance appraisal process operates. According to Murphy and Cleveland (as cited in Tziner, Murphy, Cleveland, Yavo & Hayoon, 2008), these contextual influences on raters' performance can be classified into three major categories, namely rater beliefs and attitudes toward the performance appraisal system, their orientation toward performance appraisal, and their personality (dispositional characteristics). To understand the rating process, all these types of extraneous influences must be considered.

From the above, it is clear that the rater stands central to the rating process. Rater performance – although not always sufficiently acknowledged in practice and research – is an important factor for the successful measurement and management of performance. For instance, research shows that raters have substantial limitations in information processing capability (Noe et al., 2006). Because humans have limited capabilities to process social information, we tend to make use of judgement “heuristics” or simplifying mechanisms, to make judgements about other persons (Taylor & Fiske, 1991). These heuristics – because they act as cognitive “shortcuts” that serve to simplify information processing and decision-making – potentially lead to rater errors. On the one hand, when focusing specifically on the rating of job performance, some critical authors have gone so far as to suggest that abstract measures of job performance, because of the multitude of influences that make them inherently subjective, are practically worthless (Feldman, 1994). On the other hand, all aspects of the meaning of *performance* cannot be captured sufficiently by purely objective or hard criterion measures, such as output, sales volume, or the percentage of goods that meet specifications (Feldman, 1994).

Since performance appraisal is inherently subjective, it follows that judgement plays a major role in performance appraisal (Feldman, 1994). To use a performance measure in a way in which one can extract useful information from it, specific, comparative gauges plus an understanding of the relevant context (Behn, 2003) are needed. The most commonly used appraisal techniques include: 360-degree appraisal (or multisource performance ratings), peer review, self-review, essay appraisal, graphic rating scale, field review, forced-choice rating, critical incident appraisal, management-by-objectives approach, work-standards approach, ranking methods, and assessment centres. Each of these methods has its own combination of strengths and weaknesses, and no single one is able to achieve all the purposes for which management performance appraisal systems are intended (Jordan, 2002).

Questions about performance appraisal data quality are a major component of the general “criterion problem”. In a typical research study, the criterion measure is the outcome measure (or dependant variable). This criterion problem refers to the specification and measurement of job or task behaviour, and often its evaluation on some level of quality or benefit to the organization (Feldman, 1994). These measures then serve as criteria for the evaluation of selection strategies, training programmes, motivational interventions, and so on, as well as input to a wide variety of personnel decisions (Feldman, 1994). In criterion-related validation, the criterion should be reliable, valid and free from third-variable biases, but a major question in construct validation is whether extraneous – or job irrelevant – sources of variance influence the measures (Guion, 1998). As suggested above, research has evolved to consider the influence of cognitive processes of raters as well as social and contextual variables influencing performance appraisal measures.

More than the financial and morale implications of an effective and accurate performance management system, the adequacy of a performance measurement component hold legal implications. With new legislation concerning labour relations, employment equity, and the Constitution, the possibility of a legal review of termination, promotions, pay decisions and other HR issues is a reality in South Africa. For example, the Labour Relations Act, 1995 (Act No.66 of 1995, Republic of South Africa, 1995) stipulates that when considering a dismissal, it must be both procedurally and substantively fair. In other words, when dismissing an employee on grounds of poor work performance (one of the legal reasons according to the Act), evidence about the quality of the performance appraisal system in the company will be vital. Therefore, this process will have to be legally sound to avoid liability.

Unfair discrimination is further prohibited by the Employment Equity Act, 1998 (Act No. 55 of 1998) by stating that:

No person may unfairly discriminate, directly or indirectly, against an employee, in any employment policy or practice, on one or more grounds, including gender, sex, pregnancy, marital status, family responsibility, ethnic or social origin, colour, sexual orientation, age, disability, religion, HIV status, conscience, belief, political opinion, culture, language and birth.

Discrimination is not prohibited per se, as it in essence only means making distinctions based on certain factors. Fair discrimination distinguishes those highly likely from those less likely to achieve a performance standard (Guion, 1998). What is prohibited, though, is discrimination which is done unfairly. Unfair discrimination exists when persons with equal probabilities of success on the job have unequal probabilities of being hired for the job



(Guion, 1998). Guion states that the analysis of bias attempts to ferret out potential instances of unfair discrimination. He also observes that research during the past 25 years has focused mainly on statistical fairness models and demographic moderators when studying bias. He suggests that one should seriously study rating error, as even presumably internally consistent measures are often multidimensional with several systematic contaminating influences. He further proposes that a new research agenda should abandon the idea that all variance not accounted for by a major factor is due to random error; it should call for programmatic studies of the sources of systematic error.

Only more recently, performance appraisal models have been viewed as the product of a set of social cognitive operations, which includes acquisition of information through the observation of performance, organisation and storage of that information in memory, retrieval of information from memory, and its integration to form a judgement (DeNisi, Cafferty & Meglino, 1984). Ilgen, Barnes-Farrell and McKellin (1993) recommend more attention be given to the content of cognitive variables in performance rating. The emphasis on the content of cognitive variables suggests a more integrated, multivariate approach to research on ratings, for example using path models. Cognitive processes must be included, but so must rater characteristics, ratee characteristics (including those to be rated and likely contaminants), situational variables, and the content of formal and informal procedures followed when providing the ratings (Cascio & Aguinis, 2007).

Rater source factors are understood as one of the key factors that affect rating data quality. Rating source substantially influences rating score variance (Hoffman, Lance, Bynam & Gentry, 2010). Since rating source has been shown to affect ratings beyond trait-related factors, a growing research stream has sought to investigate the possibility that similarity

between the rater and ratee affects scores that are produced in the rating process (Cascio & Aguinis, 2007; Guion, 1998). Research in related subjective interpersonal assessment contexts – like interviews and assessment centre rating – has found inconsistent support for the notion that demographic similarity between raters and ratees affect rating score variance (e.g., Sacco, Scheu, Ryan & Schmitt, 2003, and others reviewed later). Also in the performance rating domain, some research has investigated the effect of similarity on ratings, but these studies have mostly relied on similarity measures based on physical similarity characteristics, like demographic similarity (Strauss et al., 2001). Even prior to 1980, when individual differences in raters were the focus of many rating research studies, similarity research tended to focus on second-level/demographic differences (e.g., gender, ethnicity or work experience) rather than first-level direct influences such as cognitive operations or feelings toward the rating (Landy & Farr, 1980).

It is possible that the inconsistent nature of the findings about similarity effects based on “hard” characteristics like demography, oversimplify the complexity of similarity in interpersonal judgment. In the social-cognition literature, substantial evidence exist that acquaintance between the judge and the target of rating shifts the focus of similarity judgment to “deeper” – sometimes unobservable – characteristics, like values, motives and attitudes (Taylor & Fiske, 1991). When considering the nature of performance rating, where raters and ratees spend more time together than is typical in interviews or assessment centre contexts – interviews and assessment centres present raters with only brief exposures to limited behavioural information – it is possible that this level of increased acquaintance between judge and target makes deeper-level characteristics more salient as objects of similarity judgment.

Many theorists have recognized, in particular, the important role that personality similarity plays in organisational behaviours (Antonioni & Park, 2001) and even though research on the connection between rater-ratee personality similarity and ratings is quite sparse, research on the similarity effect has been identified as a fruitful area for performance ratings research quite some time ago by the seminal work of Landy and Farr (1980).

It has been suggested that a good amount of unexplained variance in ratings may be explained by rater-ratee dyadic relationships (Wexley & Klimoski, as cited in Antonioni & Park, 2001). For this reason, the present study investigates the possibility that similarity between raters and ratees, based on non-demographic characteristics such as personality, systematically affects ratings of task performance assigned by ratees. This research holds potential benefits to both theory and practice. Theoretically, it is possible that similarity research has stalled due to an inadvertent focus on demographic characteristics which essentially denies the complexity of human person perception in the workplace. If personality similarity effects are found in the study proposed, it could suggest more research that explores non-demographic variables in similarity research. Practically, the study could enhance our understanding of unwanted sources of variance in performance ratings. In this way, the study could suggest ways in which future research could seek to minimise these extraneous effects on performance rating by means of an appropriate rating system design, rater training, rating procedure, or even statistical methods to control for these unwanted effects. Last, the practical value of the current study could be beneficial to any task performance rating context, despite the fact that a specific context (higher education) was used as a result of a sample of convenience drawn from this population. Despite the present study's intent to generalise its findings to general workplace rating contexts, understanding the effect of rater-ratee similarity on lecturer ratings within the higher education context is

also seen as a potentially fruitful contribution to higher education performance management practice.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

The literature review begins with a discussion of the performance appraisal process in general and will consider some models of the process, which were developed during the past few decades. Performance appraisal problems are then discussed according to three categories, namely the process and format, the ratee's role, and the rater's role. The effect of generalizability of earlier research is also considered. After this, measurement error in performance measures followed by, more specifically, rater bias and psychometric research on ratings of performance are discussed. The researcher then states the theoretical base and mechanisms that underlie the concept of similarity effects followed by general research findings on similarity effects. For the more specific purpose of the proposed research, personality as a construct is then explored together with previous studies on personality similarity. The reason for choosing personality as similarity variable is provided. Lastly, the literature study concludes with the formulation of the research question, objectives and hypotheses.

#### **2.2 Performance measurement in general**

Performance measures in general can be defined as “a metric used to quantify the efficiency and/or effectiveness of an action” (Neely, Gregory & Platts, 2005, p. 1229). There are different ways of judging someone's work performance, for instance ranking, pair comparison estimation, and other forms of worker-to-worker comparison. Performance rating

(or performance appraisal) is one particular form of performance judgement (Landy & Farr, 1980). The importance of performance appraisal has been questioned, but the benefits outweigh the shortcomings (Strauss et al., 2001). There are many different methods and formats for conducting these ratings and the results can be used as criteria or predictors (Guion, 1998). In the history of performance appraisal, research has mostly been guided by questions about the form and content of rating scales, and how form and content relate to the adequacy of the resulting ratings as indicators of the construct of performance (Landy & Farr, as cited in Feldman, 1994). The method of rating selected should fit the specific needs of the situation. Levy and Williams (2004) argue that identifying, measuring, and defining the organisational context in which appraisal takes place is an essential part to truly understanding and developing effective performance appraisal systems.

Appraisal comprises the interaction between two or more people and immediately includes the mystery of human behaviour in the process. Even though the process has been researched, there is always room for improvement. The challenge to the researcher and the practitioner is to pinpoint areas where bias occurs and to develop ways to reduce the probability of such bias in judgement and subsequent rating processes. Measurement of performance remains complex, frustrating, difficult, challenging, important, abused and misused (Sink, as cited in Ledbas, 1995).

For the purpose of this study the appraisal process should be considered from the appraiser or rater's perspective. Appraising performance from the rater's perspective is understood as a process of cognitively processing information in order to make judgements/evaluations that are associated with others (Ilgen et al., 1993). A few of these models are considered.

## **2.2.1 Performance Appraisal Models**

Various performance appraisal models have been proposed to clarify the process of appraising performance. Feldman (1994) mentions that research in attribution theory tradition tried to explain performance judgements in terms of causal attributions for observed performance based on immediately available stimulus information. These approaches could not, however, explain the presence of both rating error and valid judgement outcome in a single process. Nor could they explain how the appraisal process took place in the real world. Feldman (1981) then provided a comprehensive account for the appraisal process to include interpersonal and organisational factors influencing both judgements and rating responses. Furthermore, DeNisi et al. (1984) presented a model which was based on attribution theory as well as the representation of person judgements and focused on the perceiver as an active information gatherer. A short discussion of these two models follows below.

### **2.2.1.1 Model proposed by Feldman**

The following model proposed by Feldman (1981) included interpersonal and organisational factors, which influence both judgement and rating responses. Attention, categorisation, recall, and information integration are the phases carried out via an automatic or controlled process (Feldman, 1981). He later distinguished between the automatic and controlled processes, by saying the “automatic” categorisation corresponds to “perceptual” mode, and the “controlled” to “inferential” mode (Feldman, 1994). Furthermore, he said that the automatic process was dominant except when decisions were problematic, in which case a consciously monitored categorisation process took place. All of these phases are, however, vulnerable to bias.

Bias can, for instance, occur during the categorisation process as performance evaluation is effected by limiting and selecting information about the employee/ratee, because memory-based judgements are made and stimulus-based judgements are influenced through the operation of attributional bias (Feldman, 1981). During the recall phase, the subsequent recall of an employee is biased by the attributes of prototypes, which represent categories to which an employee/ratee has been assigned. Furthermore, he showed that dispositional and contextual factors influenced the availability of categories during the assignment as well as the recall phase. Under some circumstances, integration of information to form beliefs or evaluations is unnecessary and is simply recalled as stored output of earlier integration. When cognitive integration is necessary, however, the controlled attribution process will start, and both causal and dispositional attributions will result (Feldman, 1981).

#### **2.2.1.2 Model proposed by DeNisi, Cafferty and Meglino**

Later, DeNisi et al. (1984) proposed a model, which reflected the view that performance appraisal was an exercise in social perception and cognition embedded in an organisational context, which require both formal as well as implicit judgement. This model emphasizes different aspects of the appraisal process than that of Feldman in 1981 (Feldman, 1994). The model consists of three phases, namely the observation phase, the information gathering phase and the integration phase and is depicted in figure 2.1. At any one of the steps, a number of cognitive sub processes operate, which are affected by a variety of personal and contextual factors (DeNisi et al., 1984).



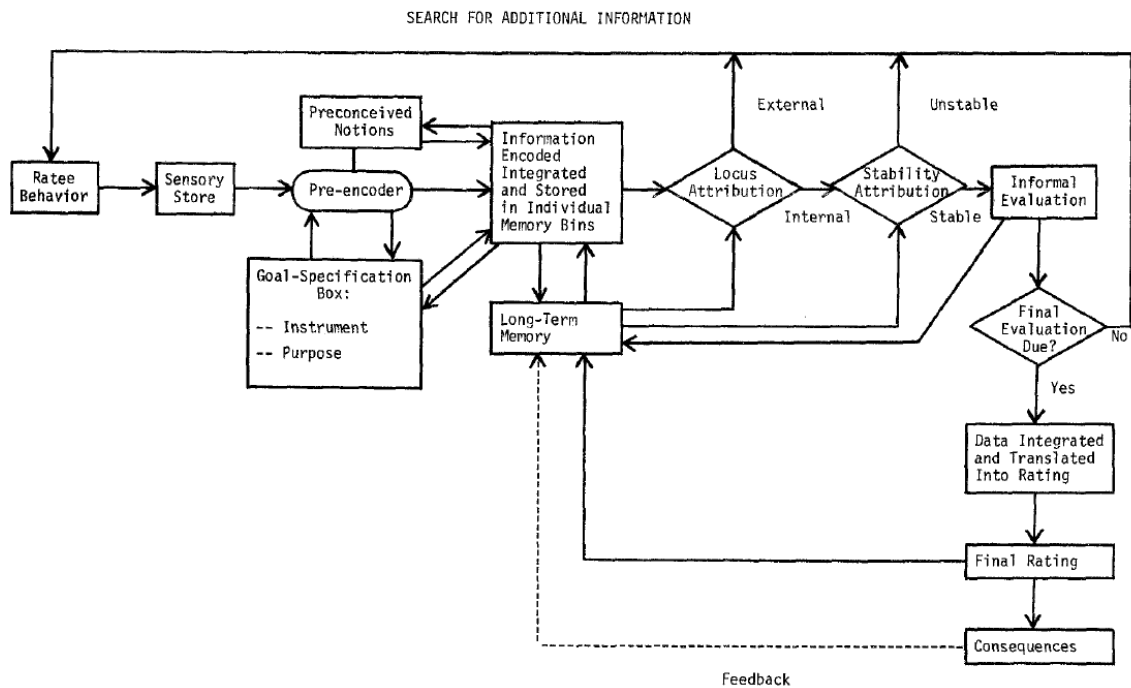


FIG. 1. Proposed model of the appraisal process.

**Figure 2.1. Proposed model of the appraisal process (DeNisi et al., 1984)**

When affected by personal and contextual factors the performance appraisal process becomes complex and open to bias during the observation phase already (DeNisi et al., 1984). For example, what is “observed” is often not behaviour per se, but rather the product of behaviour, such as written reports, customer complaints or units produced. They further argued that how a rater searched for information would determine what behaviour was observed by the rater. Four potential determinants that may influence the type of information sought are proposed: (1) preconceived notions that the rater has about the ratee; (2) the purpose for which the appraisal is conducted; (3) the nature of the rating instrument used; and (4) the time pressure operating on the rater.

During the next phase, the information gathering phase, information is encoded and stored in memory for later retrieval (DeNisi et al., 1984). This means that the rater does not store raw information, but first interprets it, and then stores the interpreted representation. The way in which information is interpreted has major consequences for the ultimate use of the information. Also, the form and extent of information processing depends on the motivation an individual has to devote cognitive resources to a problem, and the individual's capacity (given the limited resources) to process information at the same time of decision, judgement, or perception (Feldman, 1994).

The final phase of the model entails the integration of information retrieved from memory into a single evaluation, which is converted into a rating on a scale (DeNisi et al., 1984). This phase leaves room for many rating inaccuracies, because of improper weighting and combining of information.

### **2.2.1.3 Summary of performance appraisal models**

In summary, both these models are concerned with the accessibility and diagnosticity of information used to make performance judgements, and with situational and individual difference factors influencing these judgements (Feldman, 1994). Both models further agree on the importance of encoding, storage, retrieval, and integration processes and the factors that influence them. These phases are also supported by Guion (1998) who similarly indicates that ratings require at least three things: (1) a source of information, preferably observation or records, (2) organizing and remembering that information in preparation for rating, and (3) quantitatively evaluating what was remembered according to some rule. Ilgen et al. (1993) conducted a review of research on the performance appraisal process during the 1980's. They

concluded that the appraisal process includes three phases, which are: (1) acquisition of information about ratees, (2) organisation and storage of this information in memory, and (3) recall and integration of the information in such a way that it leads to a recording of an evaluation of the ratee. It seems as if some consensus exists with regard to the phases that play a part during performance appraisal. Furthermore, these phases are assumed to be loosely temporally related.

Despite the observation that appraisal takes place in phases, the end product of performance appraisal is a rating of one individual as produced by another (Ilgen et al., 1993). There are, however, a wide variety of factors that influence this rating recorded on an appraisal form (Ilgen et al., 1993). It is further mentioned that research on the appraisal process focused on the judgement component of rating, by choice. The direct result of the role of judgement in the performance appraisal process is that empirical research in this area is dominated by concerns for performance appraisal accuracy as the primary criterion of interest. This leads to the next section where possible problems with regard to performance appraisal are explored.

### **2.2.2 Performance appraisal problems**

Performance appraisal itself, as well as the process thereof, is certainly not without flaws (Von Elverfeldt, 2005). These problems can be categorised into three areas: (1) the process and format, (2) ratee's role and (3) rater's role (Kondrasuk, as cited in Von Elverfeldt, 2005).

#### *(1) The process and format*

The context in which rating occurs has an effect on various psychometric properties of the ratings (Landy & Farr, 1980). For example, studies have shown that ratings are more lenient

under conditions of administrative use than under conditions of research use (Landy & Farr, 1980). Hollander (as cited in Landy & Farr, 1980), on the other hand found no difference between administrative and research conditions in terms of reliability or validity of ratings. Davis and Landa (1999) found that the absence of fair procedures increases distress because the results of the performance appraisal are out of the control of the ratee. However, if ratees are confident in the fairness of performance appraisal process, they are more likely to accept ratings, even adverse ones (Roberts, 2003). Thus, the employees' attitudes toward the system will have a role in the prediction of willingness to meet the goals expected of them (Harris, 1988).

## *(2) Ratee's role*

The effect of the ratee's characteristics and role will be discussed briefly with regard to each of the performance appraisal phases.

Attention and observation – During this phase, Ilgen et al. (1993) noticed that raters would spend significantly less time observing ratees who were more easily stereotyped or classified than those who were less easily classified. Their finding is supported by earlier social perception literature, which found that the confirmation of a prototype leads raters to believe that they had sufficient information about the person being rated, thereby making it unnecessary to collect additional information (Feldman, 1981). Also, more time spent observing the ratee increases the accuracy of the rating (Favero & Ilgen, 1989). Finally, Balzar (1986) examined the effects of initial impression on the recording of performance information. He found that initial impression, whether positive or negative, produced a significant contrast effect, which biased later ratings. This shows that sometimes evaluations are formed prior to the gathering of information, which, in turn, affect what is observed.

Categorisation and storage process – A study conducted by Foti and Lord (1987) suggested that using prototypes to store observations may interfere with accurate recall, particularly for those behaviours related directly to the prototype.

Recall and evaluation – Early work on performance appraisals focused on trait-like characteristics of ratees (Ilgen et al., 1993). Schmitt and Lippin (1980), for example, investigated the effect of rater and ratee race and gender on rater accuracy with the outcome that raters rated same race ratees' performance more accurately. Ratee characteristics have demonstrated an effect on rating accuracy over a wide range of constructs, including specific individual difference characteristics and also consistency with stereotypes and the pattern of behaviour over time (Ilgen et al., 1993). The variance accounted for by ratee characteristics was not very high in any of these cases (Ilgen et al., 1993). They suggest that a reason for this may be that the sum total of ratee effects are likely to be underestimated since no study has investigated more than one or two ratee characteristics at a time.

### *(3) Rater's role*

The focus on raters has dominated the literature on performance appraisal since 1980 and there is a wide variety of factors that influence the rating recorded on an appraisal form (Ilgen et al., 1993). The effect of the rater's characteristics and role will also be discussed briefly with regard to each of the performance appraisal phases.

Attention and observation – The quality of ratings depend, at least in part, on the quality of observations. Raters tend to observe selectively; they attend to some behaviour, but not all. Furthermore, impressions formed by raters have strong a effect on ratings and may even have more influence than actual observation. Impressions are as likely to be formed from

stereotypes or prototypes as from direct observation, although direct observations may be distorted because of stereotypes or prototypes.

Implicit personality or implicit performance theories have a strong effect on the rater's observations and on the factor structure of the resultant ratings (Johnson, 1994). A study conducted by Foti and Lord (1987) suggested that implicit personality theories guided what was observed and recalled when raters had no knowledge of the tasks and goals of the individuals being rated. It was also mentioned that raters were extremely limited in their abilities to provide accurate ratings (Foti & Lord, 1987). This is important to remember as most people seem to expect too much from the rating system.

Considering the individual difference characteristics that a rater brings to an evaluation may have an influence on the cues that he or she perceives to be present when observing a ratee's behaviour (Ilgen et al., 1993). Observations may also be influenced by preconceived ideas that a rater has about what a ratee should do. It is therefore important for raters to have knowledge about the tasks/goals of their ratees.

Research conducted by Foti and Lord (1987) indicated that people observed and recalled most accurately behaviours which they were cued to attend to. To conclude, one could say that what is observed affects appraisal accuracy. Observation effects are often not measured directly, but rather inferred from recall (Ilgen et al, 1993). This inference is always open to the opposite explanation that recall was influenced by memory rather than observation.

Categorisation and storage process – Because performance appraisals rarely occur directly after the observation of performance takes place, it happens that the rater categorises and

stores information until recall is required (Ilgen et al., 1993). Because this process may take up to a year, it may result in some information being lost, discarded or distorted before being recalled for the appraisal. During the categorisation and storage phase, the intended purpose of the raters' observations and the amount of information they possessed about the goals of an event influences how raters categorise their observations (Foti and Lord, as cited in Ilgen et al., 1993).

There are different models describing the processes used by raters to store information. Nathan and Lord (as cited in Ilgen et al., 1993) compared the so-called traditional model by Borman and the cognitive categorisation model described by Feldman (1981). According to the traditional model, raters store information in relatively independent dimensions, whereas according to the cognitive categorisation model, information is stored more globally as a general impression. This could, for instance, have an influence on halo error.

Recall and evaluation – A number of studies have investigated the effects of rater characteristics on rating traits. For example, the study by Wexley and Youtz (as cited in Ilgen et al., 1993) showed that raters' beliefs about human nature were associated with rating accuracy. Also cognitive skills and traits, such as cognitive complexity, have been suggested as characteristics that should influence raters' ability to accurately differentiate among employee strengths and weaknesses (Smither and Reily as cited in Ilgen et al., 1993). Lastly, raters seem to be more lenient in “for real” ratings than in making ratings for research use only (Guion, 1998). Evaluation and memory are so intertwined that it is difficult to separate them under the typical conditions in which performance appraisal takes place (Ilgen et al., 1993).

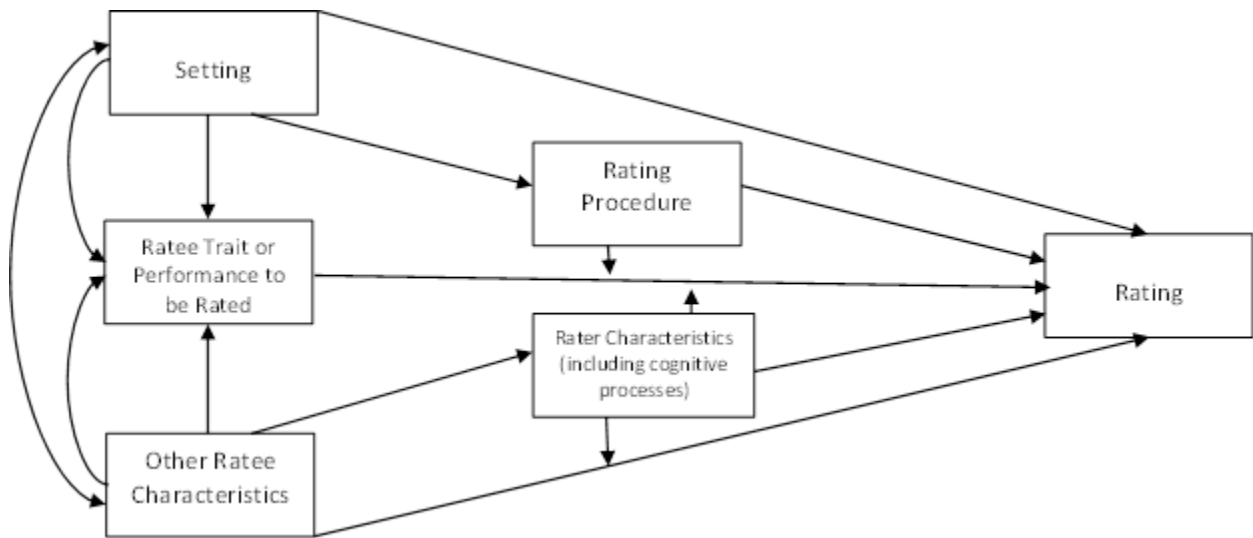
According to Kondrasuck (as cited in von Elverfeldt, 2005) problems concerning the rater's role emerge in particular because of conflicting roles of being coach and judge at the same time, lack of rater training or idiosyncratic biases as favouritism, subjectivity or leniency. Here one can see the importance of an objective rater for any effective performance appraisal, but raters are likely to make unintentional (or intentional) mistakes when assigning performance scores.

### **2.2.3 Generalizability**

The path model in figure 2.2 (from Guion, 1998) suggests classes of variables that might affect ratings, whatever their nature.

As one can see, ratings can be influenced, even if they should not be, by ratee characteristics other than those to be rated. This includes, but is not limited to: demographic characteristics, general mental ability, street smarts, or personality. Rater characteristics may also include all of the above-mentioned and also the rater's understanding of the characteristic rated, knowledge of the rating procedures, prior impressions of ratee characteristics, prejudices or response biases, general intelligence, or ability to do whatever behaviour is rated or produces the product rated. Lastly, interaction of rater and ratee characteristics may provide its own source of variance.





**Figure 2.2 Path model of classes of variables affecting ratings**

(Guion, 1998)

With so many potential sources of variance added to usual sources of random error variance, the logic of generalizability analysis seems essential in the psychometric evaluation of ratings (Guion, 1998). Guion (1998) further states that when more than one rater is required, they are frequently not independent from one another and so good generalizability across raters can be an artefact of common information about ratees.

As figure 2.2 shows, ratings are influenced by many sources and controlling nearly all of them leaves little room for strong effects of the variables studied. Also, in field studies, with lots of environmental factors as well as rater characteristics uncontrolled, weak effects are further weakened. Guion (1998) concluded that recommendations to improve ratings will not differ much from advice given prior to 1980: know the job, the standards of performance, and the reason for rating; observe often; and focus on the purpose so that extraneous considerations have minimal effect on ratings. In spite of this advice, research continues to try and identify and minimise measurement errors in performance appraisal.

### 2.3 Measurement Error in Performance Measures

Measurement error, or bias, refers to systematic group differences in item responses, test scores, or other assessments for reasons unrelated to the trait being assessed (Guion, 1998). Feldman defines bias as, “differences in the judgements of persons (or their performance or behaviour) associated with membership in a specific socially defined category”. Furthermore, Guion (1998) states that rater characteristics not being rated are sources of bias if they influence the ratings; they basically reduce validity. Bias can therefore be viewed as invalidity, which is something one would definitely prefer to avoid in any measure.

There are many different sources of bias to consider within the rating context, as is evident from the discussion of the performance appraisal models. One possible source of bias is the matter of how well rater and ratee get along on work matters with each other (Duarte, Goodson & Klich, as cited in Guion, 1998). Another source is the “same as me effect”, which will be explained below. Although similarity of rater and ratee may be a source of bias, it should not be automatically assumed without considering alternatives. For example in a study conducted by Lin, Dobbins and Farr (1992), ratings of interviewees had a small, but significant “same-as-me” effect for race. Their alternative interpretation, however, was that minority candidates may have been more comfortable in a same race setting and therefore gave more information to the interviewer.

One can see that bias generally exist easily in the performance measurement process as this process consists of a few phases. Our interest is more specifically in rater bias and trying to understand their behaviour with regards to the rating of performance.

### 2.3.1 Rater bias in Performance Measurement

This broad category of effects, called rater bias, refers to the systematic variance in performance ratings that is associated in some way with the rater and not with the actual performance of the ratee (Scullen et al., 2000). Early research on rater characteristics already pointed out different aspects which influence the rater when conducting a performance appraisal although it provided relatively few general conclusions (Landy & Farr, 1980). Wherry's theory of rating indicates that there are three broad types of factors that influence the rating of performance: (1) the ratee's actual job performance, (2) various rater biases in the perception and recall of that performance, and (3) measurement error (Scullen et al., 2000). In this section the focus will be on the various rater biases.

Scullen et al. (2000) distinguish between two major types of rater bias effects. The first is idiosyncratic tendencies and the second is associated with the raters' organisational perspective. *Idiosyncratic tendencies* shown by individual raters are also known as judgemental biases which result from some systematic measurement error on the part of the rater (Cascio & Aguinis, 2005). This includes several types of effects, for example halo, leniency/severity and central tendency. Halo refers to the tendency of raters to assign ratings for different dimensions based on an overall general impression (Scullen et al., 2000). Leniency refers to the tendency to give extreme ratings and central tendency to over confidence on the middle of the scale (Scullen et al., 2000). Studies which examined idiosyncratic rating variance found its effects to be significant (Conway, 1996; Viswesvaran, Ones & Schmidt, 1996).

Consideration of the bias which relates to the *rater's organisational perspective* include the rater's perspective toward self, subordinate, peer, and/or boss. According to Scullen et al. (2000), several researchers argued that a rater's perspective might influence performance ratings independently of the idiosyncratic tendencies described. Borman (1997) formulated three reasons to explain why one could hypothesise that perspective-related biases affected performance ratings. Firstly, raters from different organisational perspectives may focus their attention on different aspects of the ratee's performance. Secondly, raters from different perspectives may concentrate on the same aspects of performance, but attach different weights to them. Thirdly, raters from different perspectives often observe different samples of a ratee's behaviour. However, in a study conducted by Mount, Judge, Scullen, Sytsma and Hezlett (1998) where they simultaneously examined both idiosyncratic and perspective-related effects, they found that idiosyncratic effects were stronger than perspective-related effects.

There are existing training programmes to minimise these biases. According to Cascio and Aguinis (2005) the three broad objectives covered in training programmes to minimise these rater biases are: (1) to improve the observational skills of raters by teaching them *what* to attend to, (2) to improve the ability of raters to communicate performance information to ratees in an objective and constructive manner, and (3) to reduce or eliminate judgemental biases. Research has confirmed that rater training is well worth the effort (Cascio & Aguinis, 2005) and it has generally been shown to be effective in reducing rating errors, especially if the training was extensive and allows for rater practice (Landy & Farr, 1980).

As mentioned earlier, the classical psychometric errors made by raters are central tendency, leniency/severity and halo effect. Other influences on errors may be prior information about a

ratee, or prior impression based on knowledge of prior ratings (Guion, 1998). He also mentions individual differences influencing the ability to rate, which include rater qualifications, training, organisational level, rater traits and rater motivation.

### **2.3.2 Psychometric research on ratings of performance**

Wherry developed a theory for rating in which he noted that rating accuracy was dependent on (a) the performance of the ratee, (b) the observation or perception of the ratee's performance by the rater, and (c) the recall of the observation of performance by the rater (Wherry & Bartlett, 1982). Each of these components can be divided into subcomponents. He further assumed (based on classical mental test theory) that each component had a systematic part and a random part. The systematic part was divided into a true aspect and a bias aspect, where bias was systematic.

Regardless of the purpose or quality, ratings are measures (Guion, 1998). The word "measurement" implies there are individual differences in the trait measured, and they imply variance and the evaluation of possible sources of variance in the resulting measures (Guion, 1998). Furthermore, when considering variance in performance ratings (or "scores"), it should mainly be associated with variance in the *actual* performance of ratees. The variance, however, stems from influences of the measurement procedure, irrelevant worker characteristics, characteristics of the situation in which performance is measured, and characteristics of the raters. In summary, common psychometric problems are intensified in ratings (Guion, 1998).

## **2.4 Theories underlying similarity effects**

There are certain theories that underlie the effects of similarity. Landy and Farr (1980) examined the research literature that investigated whether certain combinations of rater and ratee characteristics affected performance ratings. They concluded that in one sense, rater and ratee characteristics were fixed and that they could not be as easily changed as the format of a rating scale. On the other hand, there has been some indication that training in the use of a particular rating format is of value in reducing common rating errors. There is a complicated interaction between interpersonal similarity and performance ratings (Strauss et al., 2001). Selected theories explaining how similarity influences ratings will be discussed briefly. These theories include social identity theory and the ‘similar-to-me’ hypothesis mentioned earlier, which is supported by self-categorization theory and similarity-attraction hypothesis.

The following section provides insight into the prevalence of similarity effects as suggested by an aggregation of psychological theories. Social Identity Theory is the “core” of this section, with many overlapping and corresponding insights from similar psychological perspectives.

### **2.4.1 Social Identity Theory (SIT)**

People organize complex worlds by organizing information, classifying people and judging situations and decisions according to their cognitive ability. When a situation presents itself (an event, person or other stimuli), individuals use formed schemas to make sense thereof and categorize the stimuli “appropriately” (Sacco et al., 2003). They propose that these schemas change over time as life is experienced more thoroughly and specifically. In a judgement

context, these schemas are unconsciously used to categorize the judgement outcome in terms of a perception a rater holds of the ratees, for example, demographical group or context. It is important to note that these schemas tend to change and adapt as the context one find oneself in changes (Barlsalou, 1982 as cited in Sacco et al., 2003).

The overarching theory holding the above-mentioned ideas together is the Social Identity Theory (SIT) as conceptualised by Henry Tajfel (Tajfel & Turner, 1986). SIT is concerned with group situations and starts from the assumption that social identity is derived primarily from group memberships (Brown, 2000). It further proposes that people aim to achieve or maintain a positive social identity and that this positive identity derives largely from favourable comparisons that can be made between the in-group and relevant out-groups (Brown, 2000). Should one's identity be unsatisfactory, people may plan to leave their group or find ways of achieving more positive uniqueness for it. Adding to these basic principles, Tajfel and Turner (1986) also noted that there were three classes of variables that might influence intergroup differentiation: people must be subjectively identified with their in-group; the situation should permit evaluative intergroup comparisons; the out-group must be sufficiently comparable and that pressures for distinctiveness should increase with comparability. Because of the inconsistent way humans attach value to the same constructs, people and events, SIT introduces the possibility of similarity bias in judgement contexts – highlighting the possibility of one individual judging another through a lens of personal social identity that disturbs an objective and mechanical view to some extent.

#### **2.4.2 The mechanisms of similarity effects**

There are different mechanisms of similarity effects. According to Antonioni and Park (2001), previous research identified two theories which explain how similarity influences rating. One is based on the similarity-attraction paradigm and the other on leader-member exchange theory (Antonioni & Park, 2001).

The idea of similarity-attraction has been established firmly by social psychology through proving the positive relationship between interpersonal affect and similarity between individuals on various dimensions, for example attitudes, personality, and demographic characteristics (Antonioni & Park, 2001). The similarity-attraction hypothesis borrows from learning theory by using a reinforcement framework to explain why similarity affects people's evaluation of others (Strauss et al., 2001). The model views evaluative responses as a function of reinforcing stimuli associated with conditioned stimuli. For example, similar attitudes are perceived as being rewarding and are viewed in the model as positive reinforcements, whereas dissimilar attitudes will be seen as negative reinforcements. According to the model, an affective response (e.g. interpersonal attraction) mediates the relationship between the conditioned stimulus (e.g. similarity) and evaluative response (e.g. performance rating). One can conclude that similarity is a reinforcing stimuli leading to an evaluative response. In addition, previous studies have also shown that interpersonal affect is significantly associated with ratings of leadership and management behaviours (Antonioni, 1999), work behaviours, or ratings on annual performance appraisals (Varma, DeNisi & Peters, 1996). Furthermore, interpersonal affect could influence ratings directly or through various cognitive processes, for example by influencing attention attribution, retrieval or integration which are cognitive phases of the models discussed earlier.



Concomitant with the similarity-attraction idea is the self-categorization theory, conceptualised by Turner, which argues that our self-concept is based on the social categories we place ourselves in (e.g. age, gender, and race) and that we have the desire for a positive self-identity (Strauss et al., 2001). The need to have a positive self-identity leads to the preference for and more positive evaluation of those similar to us on the social categories on which we base our identity. This theory further suggests that the categorisation of others do not necessarily require interaction with them, but merely a glance (Strauss et. al., 2001).

The other theory explaining how similarity influences rating is different as it proposes that interpersonal similarity may actually influence how people behave at work (Antonioni & Park, 2001). This is the leader-member exchange theory. The theory conceptualizes leadership as a process of interaction between leader and follower and centres on the dyadic exchange relationships between both (Winkler, 2009). Similarity in leader-member dyad plays an important role in the building of trust between leaders and members through its influence on early interactions in relationships. During the early stages of a relationship, one has little information about the other person. You will therefore increase trust to those you are similar to – with whom you share similar outlooks. This initial trust may have a lasting effect in the quality of the relationship.

The above-mentioned ideas lead to the “similar-to-me” hypothesis which basically argues that people will be rated higher the more similar they are to the rater or the more similar the rater believes people are to him/herself (Strauss et al., 2001). Similarity is thus a construct in itself and will be discussed next.

### **2.4.3 Similarity as a construct**

Since the concept of similarity is a cognitive heuristic, it can be viewed as relatively flexible (Sacco et al., 2003) which suggests that it can be changed or manipulated through knowledge, by altering the context, or by drawing people's focus to different stimuli, cues or features. As mentioned earlier, research has suggested that the basis of our similarity judgements changes as we age and gain expertise. This happens due to a shift in the key features used to make the similarity judgement from one context to another (Sacco et al.).

The role of similarity in the workplace has also been studied by borrowing from cognitive psychology's research and theories. Many of these studies' focus are on supplementary person-organisation (PO) fit. This examines the extent to which similarities among workers or between an employee and the greater organisational context impacts various work and adjustment outcomes (Muchinsky & Monahan, 1987).

Relational demography theory also employs similarity as a construct. This theory examines the importance of similarity more narrowly by focusing on how people use demographic variables to assess how similar one individual is to another. As mentioned earlier, the similarity-attraction paradigm proposes that our self-concepts are formed by the groups to which we think we belong. Should this be true, people will evaluate members of their own group more positively than those of other groups to maintain a positive self-regard. In this sense, similarity is also built into the forming of the construct self-concept.

## **2.5 General research findings on similarity effects**

Performance appraisal is a complex process affected by many factors, for example, organisational, process and interpersonal barriers. Research has been conducted on a variety of similarity effects in different scenarios over the years. The focus of previous research has mostly been on demographic similarity variables and the effects of selected individual-differences variables on the ratings of performance (Cascio & Aguinis, 2005). Gender, age and race effects on ratings have enjoyed a great deal of attention. Nevertheless, idiosyncratic variance (which is variance due to the rater) has, in fact, been found to be a larger component of variance in performance ratings than the variance attributable to actual ratee performance (Scullen, Mount & Goff, 2000). Scullen et al. (2000) found rater variance to be 1.21 times larger than ratee variance for supervisory ratings, 2.08 times larger for peer ratings, and 1.86 times larger for subordinate ratings. Individual differences within a rater and ratee as well as the interaction between the two clearly affect performance ratings. In table 2.1 a summary of findings on the interaction of rater-ratee characteristics and performance ratings can be seen which is in the scope of the current research.

As is evident from table 2.1, when considering the interaction of rater-ratee gender on performance ratings, females are rated less favourable and with greater negative bias by raters who hold traditional stereotypes about women (Dobbins, Cardy & Truxillo, 1988). Also, when considering the interaction of rater-ratee race, studies show that both white and African-American raters consistently assign lower ratings to African-American ratees than to white ratees (Oppler, Campbell, Pulakos & Borman, 1992). Race effects may, however disappear when cognitive ability, education, and experience are taken into account (Waldman & Avioli, 1991).

**Table 2.1****Summary of findings on Interaction of rater-ratee characteristics and performance ratings**

<b>Gender</b>
In the context of merit pay and promotions, females are rated less favourably and with greater negative bias by raters who hold traditional stereotypes about woman (Dobbins, Cardy, & Truxillo, 1988).
<b>Race</b>
Both white and African-American raters consistently assign lower ratings to African-American ratees than to white ratees. White and African-American raters differ very little in their ratings of white ratees (Oppler et al. 1992; Sackett & DuBois, 1991). Race effects may disappear when cognitive ability, education, and experience are taken into account (Waldman & Avolio, 1991).
<b>Actual versus perceived similarity</b>
<i>Actual</i> similarity (agreement between supervisor-subordinate work-related self-descriptions) is a weak predictor of performance ratings (Wexley et al. , 1980), but <i>perceived</i> similarity is a strong predictor (Turban & Jones, 1988; Wayne & Liden, 1995).
<b>Performance attributions</b>
Age and job performance are generally unrelated (McEvoy & Cascio, 1989).
<b>Citizenship behaviours</b>
Dimension ratings of ratees with high levels of citizenship behaviors show high halo effects (Werner, 1994). Task performance and contextual performance interact in affecting reward decisions (Kiker & Motowidlo, 1999).
<b>Personality characteristics</b>
Similarity regarding conscientiousness increases ratings of contextual work behaviors, but there is no relationship for agreeableness, extraversion, neuroticism, or openness to experience (Antonioni & Park, 2001).

(from Cascio & Aguinis, 2005, p. 108)

Despite the fact that race effects were found for Oppler et al. (1992) the nature of interviews could also affect similarity effects observed. Lin, Dobbins and Farh (1992) examined the effects of interviewer and interviewee race and age similarity on interview outcomes and found that same-race effects with conventional structured interviews were stronger than with situational interviews. They found no age similarity effects in the context of interview ratings.

Sacco et al. (2003) studied the effects of race and gender similarity in one-on-one highly structured college recruiting interviews. They found no significant racial or gender similarity effects with the use of hierarchical linear modelling (as opposed to ANOVA approaches used in earlier studies). Small similarity effects were, however, found when *D*-score and analysis-of-variance (ANOVA)-based interaction approaches were conducted at the individual level of analysis.

Sacket and DuBois (1991) compared data of the effects of rater and ratee race on rating of performance from three sources: a large-scale civilian study, a large-scale military study, and a meta-analytic study. The results of the civilian and military study converged, but they differed from the meta-analytic results. The means for the civilian data and for certain dimensions of the military data clearly indicated that Black ratees consistently received lower ratings than White ratees from both White and Black raters. Also interesting is that White and Black raters differed very little in their ratings of White ratees, but differed much more in their rating of Black ratees. Black ratees received ratings from White raters that ranged from .020 to .100 of a standard deviation lower than the ratings that they received from Black raters. Contrasting to this, ratings for White ratees from Black raters ranged from .003 of a standard deviation lower to .003 of a standard deviation higher than ratings from White raters. Another meta-analysis conducted by Kraiger and Ford (1985) reported that raters gave higher ratings to ratees of their own race. A possible moderator was the salience of race; as the percentage of Blacks among ratees increased, the race effect became less pronounced.

The meta-analysis, mentioned above, conducted by Kraiger and Ford (1985), found conflicting results of whether the existence ratee race effects were related to rater race or were moderated by situational factors. They included 88 studies in their meta-analysis. The

corrected mean correlation between ratee race and ratings for White and Black raters were .183 and -.220, respectively. They concluded that the size of the ratee race effect is small for both Black and White raters, but consistent across studies.

Another study concerned with the way the gender and race of the rater and ratee influenced the level and variance of performance ratings conducted, indicated that people rate members of their own race group with more confidence than they do members of other race groups (Schmitt & Lippin, 1980). This confidence is reflected in the variance of performance ratings and the degree to which rating correlated with actual performance. During most of the above mentioned studies, a similarity variable was calculated through difference scores and profile matching.

Findings other than those of demographic similarities which concern the interaction between raters and ratees are also of interest, for example, longer relationships between a rater and ratee resulted in more accurate ratings of performance (Sundvik & Lindeman, 1998). Actual versus perceived similarity between a rater and ratee have also been studied by certain researchers. Actual similarity is a weak predictor of performance ratings (Wexley, Alexander, Greenawalt & Couch, 1980), but perceived similarity is a strong predictor (Turban & Jones, 1988; Wayne & Liden, 1995).

In spite of the widespread research that rater judgement has received, the findings seem to be inconclusive. Perhaps “softer” factors could be taken into consideration as one move away from hard demographic similarity influences. It would seem that aspects like relationships with colleagues and engagement in one’s work environment come into play. In the fast-moving and evolving business world the retention of employees is a major factor. If

performance appraisal is implemented correctly, it may influence retention. Performance appraisal can be improved by identifying and minimizing construct-irrelevant factors influencing performance ratings. Participatory performance appraisal is also an essential and proven attribute of an effective performance appraisal system (Roberts, 2003).

### **2.5.1 Choice of personality as similarity variable**

Race and/or gender have mostly been used in studies concerned with rater bias. As the results seem inconclusive, it contributed to the question whether deeper rooted variables play a role in rater bias. The psychological perspectives discussed in section 2.3 lead to a few general conclusions that could be typically used to explain possible sources of idiosyncratic rater effects. When considering all these ideas and theories, it could be argued that it is almost central to human behaviour to subconsciously (and sometimes even consciously) perceive someone similar to oneself more positively than others who is less similar to the observer. This will then automatically be a bias-enhancing factor in any judgement process, leaving rates that are less similar to the rater in an immediately less favourable position than those who are more similar. Generally speaking, it is more commonly noted that classification and association between people happens on race and/or gender level. Neighbourhoods, countries, social classes and friend-groups are for the most part racially and culturally homogenous units. Though the exception exist and is promoted in socio-political contexts, race and gender seem to be a more noticeable cause of people either regarding themselves as similar or different to those around them. However, based on the personality theories discussed above, as one has more interaction with someone, one would go beyond mere demographic variables influencing the cognitive process on which performance appraisal is based. Deeper seated variables like personality, interests, and religion may come into play. Very few studies on the

way in which personality similarity influences rater bias exist; therefore this research attempts to add to the current small body of research concerning this variable.

## **2.6 Personality**

Personality refers to those characteristics of a person that account for consistent patterns of thinking, feeling and behaving (Pervin, Cervone & John, 2005). Personality also provides a channel through which humans interact (Polk, 2006). Some time ago there was a resurgence of interest in personality in the workplace (Strauss et al., 2001), as revealed by an increase in the use of personality tests in the selection process. Research also supports the lasting nature of personality traits (Funder, 1991).

The attempt to devise a taxonomy of personality traits formed a big part of psychological research through most of the last century. After many years it was determined that a five-factor taxonomy describes the structure of personality better than any other alternative (Kroeck & Brown, 2004). The labels of the five factors, which are still used today, were developed in 1963 by Norman (Kroeck et al., 2004). They are extraversion, agreeableness, conscientiousness, neuroticism and openness to experience. The five-factor model is a comprehensive, well-researched model of personality. The five factors are shortly defined by Taylor (2004) as the following:

Extraversion – those who enjoy being around people, assertive, active and talkative;

Agreeableness – sympathetic toward others, sincere, eager to help;

Conscientiousness – person who perseveres, is responsible and organised;



Neuroticism – refers to a person’s emotional stability and general tendency to experience negative effects in response to their environment; and

Openness to Experience – Seek out novel experience and reflect on ideas, have wide range of ideas, imaginative.

Prior to hypothesizing about the relationship between personality similarity and performance ratings, the main effects for personality of the ratee will be considered. Barrick and Mount (1991) conducted a study where they found conscientiousness showed consistent relation with job performance on three different criteria ( $r$  ranges from .20 to .23) and with regard to five different occupational groups. Extraversion was a valid predictor for occupations involving social interaction, managers and sales. And both extraversion and openness were valid predictors (extraversion:  $r = .26$ ; openness:  $r = .25$ ) for the training proficiency criteria for all five of the occupational groups. Therefore, it is hypothesised in line with literature that ratee conscientiousness, extraversion and openness will be positively related to ratee performance ratings.

### **2.6.1 Personality as a similarity measure**

Several researchers have called for the investigation of the similarity effect using personality as measures of similarity (Bauer, Green & Bauer, 1996; Landy & Farr, 1980). According to Antonioni and Park (2001), many scholars have also argued that personality similarity may allow individuals to predict the behaviour of others, which makes it likely that co-workers will interpret behaviours and environmental events similarly. Personality similarity also increases the amount of communication between individuals (Engle & Lord, 1997) as one would feel more comfortable communicating with someone you expect to understand you

better. It further increases social integration in an organisation and lessens role conflict and role ambiguity (Turban & Jones, 1988). Consequently, personality similarity could actually lead to better interpersonal relationships and higher job performance.

The above-mentioned suggests that since a rater may have more positive affect toward a ratee who is similar to himself or herself, which in turn could inflate ratings, personality similarity should be positively associated with the ratings of ratee work behaviours. Also, dyads with similar personalities may work together more effectively because they trust each other more, share the same perspectives, and communicate better, whereas on the other hand, dissimilarity between two people might increase misunderstanding and reduce social integration.

Alternative explanations are suggested for the influence of personality similarity on ratings. According to Strauss et al. (2001), some argue that personality similarity leads to positive interpersonal affect, which biases the ratings upwardly, while others oppose this argument by saying that personality similarity actually does improve work behaviours through enhancing trust and shared understanding.

These two explanations have very different implications, but few studies have actually investigated the mechanisms through which rater-ratee similarity influences ratings. Bauer, Green and Bauer (1996, p. 1563) suggest that "...other personality similarity measures need to be explored. Studying the 'Big Five' measures of personality would be a good place to start".

## **2.7 Previous studies on personality similarity and their effects on performance rating**

According to Antonio and Park (2001), there have been few empirical studies of the effect of personality similarities on ratings. In the studies that were conducted, findings were inconsistent, but this may be accounted for by some of the methodological and conceptual problems that was encountered (Antonio & Park, 2001). One of the problems, for example, was the violation of the assumption of independence. This violation occurred due to the fact that in most studies each rater evaluated more than one ratee. This is likely to introduce dependencies in the data and violate the independence assumption, which is the basis of most statistical analysis. Another problem identified was errors associated with the use of human observers in the gathering of observational bias (Wildman, Erickson & Kent, 1975).

With regard to other studies conducted on the effect of personality similarities on performance rating, Landy and Farr (1980) mentioned that previous research has examined the relationship between the characteristics of the rater and various criteria of rater effectiveness. These studies were grouped into three classes: personal characteristics of the rater, type of rater vis-à-vis ratee, and rater knowledge of the ratee and the job. In Scullen et al. (2000), they indicated that idiosyncratic rater effects accounted for over half of the rating variance and the combination of general and dimensional ratee performance was less than half the size of the idiosyncratic effect. This illuminates the effect that personal characteristics of raters have on ratings.

## **2.8 Summary and conclusion**

As mentioned earlier, there have been few empirical studies of the effect of personality similarity on ratings. The most recent two studies which are of current interest were conducted by Antonioni and Park (2001) and Strauss et al. (2001). Strauss et al. (2001) conducted a study which examined the effect of personality similarity (relational and perceived) on performance ratings of peers as well as supervisors, and also the role of familiarity and liking. During the same year, Antonio et al. (2001) examined the effects of personality similarity on peer ratings of contextual work behaviours.

Strauss et al. (2001) found minimal support for a relationship between relational (actual) personality similarity and performance ratings in either dataset. However, *perceived* similarity in these same data sets related strongly to performance ratings. The fact that individuals form different perceptions about things (Landy, Barnes and Murphy, 1978) is a possible reason for the effect perceived similarity had in this research. Furthermore, the results provided moderate support for liking as a mediator of the perceived personality similarity-performance rating relationship and no support for interpersonal familiarity as a moderator of the relationship between relational personality and perceived personality similarity.

Antonioni and Park (2001) found that rater-ratee similarity in only Conscientiousness (but not in other dimensions of the Big 5 Personality traits) was positively associated with peer ratings, even after controlling for interpersonal affect. They suggested that the observed effect of personality similarity may reflect actual behavioural differences rather than biases due to the interpersonal affect.

During the past decade it appeared as if this area of enquiry was closed and the above outcomes was accepted. As mentioned earlier, research found that extraversion was a valid predictor for occupations involving social interaction, managers and sales. And both extraversion and openness were valid predictors for the training proficiency criteria for five different occupational groups. Other studies which investigated relational personality effects on organisational outcomes using personality measures measuring components of the five-factor model related to these dimensions also found significant results. For example, Bauer and Green (as cited in Strauss et al., 2001) found that actual similarity in positive affectivity (a trait similar to extraversion) related significantly to performance ratings.

A related aspect to consider is the target dimension of performance that was measured in previous research. Performance is a multidimensional construct, consisting of task-orientated work behaviours and contextual work behaviours (Borman & Motowidlo, 1997). Research shows that task performance and contextual performance contribute independently to overall performance (Motowidlo & Van Scotter, 1994). Contextual work behaviours are important as they shape the organisational, social and psychological context which is a catalyst for task activities and processes (Borman & Motowidlo, 1997). Task performance can in turn be defined as “the effectiveness with which job incumbents’ perform activities that contribute to an organisation’s technical core either directly or by implementing a part of its technological process, or indirectly by providing it with needed materials or services” (Borman & Motowidlo, 1997, p.99). Task activities that vary across different occupations are more likely to be role-prescribed and to involve cognitive ability. To differentiate between task-orientated and contextual work performance is important to determine what the focus of research will be. According to Antonioni et al. (2001), for example, similarity regarding conscientiousness

increases ratings of contextual work behaviours, but there is no relationship for agreeableness, extraversion, neuroticism, or openness to experience.

The conclusion is that research findings with regard to this topic are inconclusive and there is substantial room for more research with regard to the effect of personality similarity on the rating of performance. Through the literature study on general similarity effects on ratings it was discovered that similarity generally does lead to more favourable ratings. When, however, extending this research to personality similarity, which is deeper seated than surface-level, superficial characteristics, it complicates the research approach that is required, since not all traits are commonly associated with higher levels of job performance (Barrick & Mount, 1991). For instance, the similarity between rater and ratee in terms of an 'unfavourable' trait such as neuroticism would probably not result in more favourable ratings of performance. Based on this argument, suitable hypotheses were developed in the present study that allowed for trait level effects in addition to simpler similarity effect hypotheses.

## **2.9 Research question and objective**

### **2.9.1 Research Question**

The need for the proposed research was initiated by the overarching need to understand the systematic sources of variance in subjective ratings of task performance. The aim of this study was to investigate the role that rater and ratee personality, as well as personality similarity between rater and ratee, has on job performance ratings, specifically in the form of task-orientated work behaviours.

## 2.9.2 Research Objectives

**The research objectives are:**

- To determine whether differences exist in ratings assigned by raters (students) to ratees (lecturers) with certain personality dimensions with regard to task-orientated work behaviour.
- To explore the extent in which rater-ratee personality similarity on any of the Big Five personality traits acts as one of the possible causes of systematic error in the performance appraisal of task-orientated work behaviours.

## 2.9.3 Present study

The present study is a constructive replication of two studies, namely Antonioni et al., as well as Strauss et al., both conducted in 2001. Antonioni et al. (2001) examined the effect of personality similarity on performance ratings in two data sets (peers and supervisors). In their study, the effect of personality similarity on peer ratings of contextual work behaviours was measured. The participants in the study represented all levels of a company and they worked in four areas, namely: programming, claims, payment processing, and product management. The sample consisted of 406 raters and 396 ratees. Antonioni et al. (2001) made use of polynomial regression and traditional regression to report their results. They found that rater-ratee similarity in conscientiousness, but not in other dimensions, was positively associated with peer ratings - even after controlling for interpersonal effect.

Strauss et al. (2001) conducted an investigation of personality similarity effects (relational and perceived) on peer as well as supervisor ratings and the role of familiarity and liking. In

their hypothesis which predicted that relational personality (similarity between *actual* rater-ratee personalities) would significantly relate to performance ratings, they made use of polynomial regression to directly test similarity effects of self-other personality traits. Difference scores are often used to operationalize self-other agreement. They mention criticism of the use of D-scores: difference scores are less reliable than their components, and they often explain less variance than their components. There was, surprisingly, minimal support for a relationship between relational personality and performance ratings in either of the data sets. They further investigated the possibility of interpersonal familiarity as moderator (between relational personality and perceived personality similarities) and liking as mediator (for the perceived personality similarity-performance rating relationship).

The present study, as a constructive replication of these earlier studies, differs from earlier investigations into the effect of personality similarity on performance ratings in the following three ways:

- *Criterion measure.* Antonioni and Park's research focused on contextual work behaviours whereas the present study focuses on task-performance behaviours. Scotter (as cited in Antonioni & Park, 2001) found that task performance and contextual work behaviours contributed independently to overall performance. For this reason, it is also important to understand the possible effect of personality similarity on task-oriented work performance ratings.
- *Predictor measure set.* Antonioni and Park selected only a few personality dimensions to focus on, whereas in the present study all personality dimensions of the Big Five personality dimensions are included.



- *Direction of ratings.* Earlier studies on personality similarity assessed downward or peer ratings of performance. Since ‘upward’ ratings of seniors is becoming a more common practice, it would be important to understand the potential effect of personality similarity on upward ratings of task performance.

## **2.10 Hypotheses**

In the present study, the following hypotheses are tested.

### **2.10.1 Personality main effects**

Derived from literature (mentioned on page 38) extraversion, conscientiousness and openness (selected out of the Big Five dimensions of personality) are expected to be valid predictors of job performance in this study and are tested with the following hypotheses:

*Hypothesis 1:* There is a statistically significant positive relationship between rater extraversion and rater performance rating.

*Hypothesis 2:* There is a statistically significant positive relationship between rater conscientiousness and rater performance rating.

*Hypothesis 3:* There is a statistically significant positive relationship between rater openness and rater performance rating.

### **2.10.2 Relational similarity hypotheses**

Based on the arguments outlined above, hypotheses were formulated regarding the role of personality similarity in performance ratings. It would be an oversimplification to hypothesise that trait similarity would result in higher ratings, as is typical of demographic similarity studies, since personality traits generally differ in the direction of main effects on performance ratings (Barrick & Mount, 1991). Personality trait similarity hypotheses, therefore, need to also consider the direction of main effect of the target personality dimension typically observed. The following hypotheses were formulated to determine whether personality similarity between the rater-ratee dyad would have an effect on task orientated work behaviour ratings:

*Hypothesis 4:* Ratings of the ratee's task performance are higher when both rater and ratee have high levels of extroversion than when both rater and ratee have low levels of extroversion.

*Hypothesis 5:* Ratings of the ratee's task performance are higher when both rater and ratee have high levels of agreeableness than when both rater and ratee have low levels of agreeableness.

*Hypothesis 6:* Ratings of the ratee's task performance are higher when both rater and ratee have high levels of conscientiousness than when both rater and ratee have low levels of conscientiousness.

*Hypothesis 7:* Ratings of the ratee's task performance are higher when both rater and ratee have low levels of neuroticism than when both rater and ratee have high levels of neuroticism.

*Hypothesis 8:* Ratings of the ratee's task performance are higher when both rater and ratee have high levels of openness than when both rater and ratee have low levels of openness.

## **2.11 An overview of polynomial regression with response surface analysis**

The role of personality similarity in performance ratings is more complex due to the fact that main effects of these traits are not always in the positive direction (Barrick & Mount, 1991). For this reason, a suitable data analysis technique is required that is able to explore the nuanced nature of both rater and ratee main effects, as well as rater-ratee interaction effects, on performance ratings. Polynomial regression with response surface analysis is the statistical technique used for the present study to discover possible personality similarity effects between the rater and ratee. Consequently an overview is given as this technique is not yet widely used. Polynomial regression with response surface analysis is a sophisticated statistical approach that has become more popular in multisource feedback research, for example self-observer rating discrepancy (Shanock, Baran, Gentry, Pattison & Heggstad, 2010). This approach allows researchers to examine the extent to which combinations of two predictor variables relate to an outcome variable, particularly in the case where the discrepancy between the two predictor variables are a central consideration.

Response surface analysis is an emerging technique that can provide a nuanced view of the relationship between the combination of two predictor variables and an outcome variable by

graphing the results of polynomial regression analyses on a three-dimensional space. Response surface methodology facilitates substantive interpretation when constraints imposed by difference scores are rejected, as is usually the case (Edwards, 2002). This technique therefore has the potential to be more explanatory than difference scores or traditional moderated regression analyses.

In general this technique can be used for any situation in which researchers are interested in the way in which combinations of two predictor variables relate to an outcome. Previously a specific similarity variable was calculated for similarity studies (e.g., Bates, 2002; Schmitt, Pulakos, Nason, & Whitney, 1996; Goldberg, 2005). These variables were calculated through difference scores and profile matching, but this is not necessary when using polynomial regression as the similarity is calculated automatically and expressed by the response surface. Examples of research questions for which polynomial regression with response surface analysis could be used include how a person's actual versus desired levels of job attributes (e.g., variety, autonomy, travel, and workload) relate to satisfaction.

There are, however, a few assumptions that must be met. According to Edwards (2002), the first assumption is that the two predictor variables must be commensurate. This means that the two predictor variables must represent the same conceptual domain. The second assumption is that the predictor variables should be measured on the same numeric scale so that their degree of correspondence can be determined. Should the variables not be measured on the same scale, researchers could transform the variables to a standardized scale by placing them on a common metric. In the current study this was not necessary and the two predictor variables were measured on a 5-point Likert-type scale from 1 = "very inaccurate" to 5 = "very accurate". Lastly, all the usual assumptions for multiple regression analysis

should be met. This includes normality, linearity and homoscedasticity, which are discussed in more detail in the results chapter.

## **2.12 Conclusion**

The hypotheses were developed in an attempt to answer the research questions which resulted from the literature review. Performance measurements as a general idea was discussed with specific consideration of performance appraisal models. Problems with regard to performance appraisal were identified, while measurement error in performance measures and more specifically rater bias and psychometric research in performance measurements were discussed. The underlying theories of similarity identified were social identity theory together with mechanisms such as similarity-attraction, leader-member exchange and self-categorization which led to the “similar-to-me” hypothesis. Next, similarity was addressed as a construct, with discussion of general research findings of similarity, which leads to the choice of personality as the similarity variable as research focus. Lastly, previous studies on personality similarity and the effect on performance ratings were discussed where after the research objectives were identified and hypotheses formulated.

If similarity effects are found, as hypothesised, it would imply that they contribute to the error variance of the performance appraisal process and should be minimised to the extent that it is possible to do so. The validity of performance appraisal procedures could be improved by eliminating possible rater bias caused by similarity effects. In the next chapter the methodology that was followed in this research is discussed.

## **CHAPTER 3**

### **RESEARCH DESIGN AND METHODOLOGY**

#### **3.1 Introduction**

This chapter introduces and explains the methodology used to determine the effect that personality similarity between rater-ratee dyads have on the rating of task-orientated performance measures.

The objectives set in the preceding chapters include determining the effect that ratee extraversion, conscientiousness and openness have on the average performance rating received and the effect that the rate-ratee dyad personality similarity (of any of the Big 5 personality dimensions) has on the rating of task-orientated performance measures.

Research was undertaken to determine the most effective way of reaching these objectives and testing the subsequently formed hypotheses. In summary, this chapter provides details of the research design, statistical hypotheses, sample information, data collection and analysis techniques.

#### **3.2 Statistical hypotheses**

The substantive research hypotheses listed at the end of chapter 2, are consequently expressed as the following statistical hypotheses which were tested in this study:

$$H_{01}: \rho = 0 \text{ versus } H_{a1}: \rho > 0$$

$$H_{02}: \rho = 0 \text{ versus } H_{a2}: \rho > 0$$

$$H_{03}: \rho = 0 \text{ versus } H_{a3}: \rho > 0$$

Also the similarity hypotheses are stated statistically as the expectation that the response surface coefficient for the slope of the line of agreement ( $a_1$ ) would be significant:

$$H_{04}: a_1 = 0 \text{ versus } H_{a4}: a_1 > 0$$

$$H_{05}: a_1 = 0 \text{ versus } H_{a5}: a_1 > 0$$

$$H_{06}: a_1 = 0 \text{ versus } H_{a6}: a_1 > 0$$

$$H_{07}: a_1 = 0 \text{ versus } H_{a7}: a_1 > 0$$

$$H_{08}: a_1 = 0 \text{ versus } H_{a8}: a_1 > 0$$

### 3.3 Research Design

A research design is a general plan for implementing a research strategy (Gravetter & Forzano, 2009). Furthermore, a research design specifies whether the study involves groups or individual participants, makes comparisons within a group or between groups, and how many variables is included in the study (Gravetter & Forzano, 2009).

For this research, a correlational study was conducted. Correlational studies are used to look for relationships between variables. There are three possible results of a correlational study: a positive correlation, a negative correlation, or no correlation.

### 3.4 Sample

The sampling technique used for the present study was a non-probability convenience sampling method. This technique implies selecting cases that are the easiest to obtain for the specific study (Welman, Kruger & Mitchell, 2005) and furthermore relies on the availability of people and their willingness to participate in the research (Gravetter & Forzano, 2003). The participants were selected from final-year and postgraduate classes and their lecturers at a tertiary institution who voluntarily indicated a willingness to participate in the research.

The lecturers of the selected classes represented the ratees or targets. They were informed that information gathered would remain confidential and their names would not be mentioned in the thesis or any other publication.

Student participants took part in the research as part of the annual lecturer evaluation. The students represented the raters. The researcher selected final-year and postgraduate students for the following reasons: (1) on that level of study classes are smaller, (2) students have been in lectures for at least three years and have seen previous lecturers perform and therefore have a better idea of what to rate and evaluate, (3) students have built better relationship with lecturers, and (4) students attach more value to giving valuable feedback to lecturers on their performance.

Although the convenience sample selected represents a specific higher education environment, the focus should remain on the rating of task performance behaviour generically and not on the sample's environment.



### **3.4.1 Participants**

The initial sample in the present study comprised of final-year and post graduate students ( $N = 154$ ) from five different classes, who rated their respective lecturers. The demographic profile of the sample is presented in table 3.1. The majority of the sample (67%) was female and 87% of the sample was between the ages of 18-25 years. With regards to their academic year, 52% was in their fourth year, 28% in their fifth year and 11% in their sixth year of studying. Lastly, in terms of ethnic group, 83% of the sample was White, 15% Coloured, 1% Asian and 1% Black.

The above information shows the demographic characteristics of the sample under investigation in the present study. These demographic variables were included due to the effect it might have on the strength of the personality similarity between student and lecturer.

### **3.5 Measurement instruments**

Measurement is the assignment of numbers to attributes or properties of people, objects or events based in a set of rules (Stevens, 1968). Measurement is essential to any research as it allows the issues under investigation (Aguinis, Henle & Ostroff, 2001) to be described,

Table 3.1

*Demographic details of the sample*

Student Age		
Responses	Frequency	Percentage
18 – 25	131	85.1
26 – 30	16	10.4
30 +	4	2.6
Unreported	3	1.9
Student Gender		
Responses	Frequency	Percentage
Male	49	32.7
Female	101	65.6
Unreported	4	2.6
Ethnic Group		
Responses	Frequency	Percentage
Black	2	1.3
White	124	80.5
Coloured	22	14.3
Asian	2	1.3
Unreported	4	2.6
Academic Year		
Response	Frequency	Percentage
3rd year	13	8.2
4th year	78	49.1
5th year	43	27
6th year	17	10.7
Unreported	8	5

Degree Course		
Response	Frequency	Percentage
BComm Psych (Hons)	11	6.9
NOS/PGCE	86	54.1
BComm Marketing (Hons)	24	15.1
BComm (Management Sciences)	15	9.4
BComm (Hons) Financial Analysis	17	10.7
Unreported	6	3.8

predicted, explained, diagnosed and decided. The accuracy of this research therefore relies on the sound measurement of the variables examined.

### 3.5.1 Personality measure

The Mini-IPIP developed by Donnellan, Oswald, Baird and Lucas (2006) was used to measure the Big Five personality factors. This is a 20-item short form of the 50-item IPIP Five-Factor Model, with four items per Big Five trait. This Mini-IPIP was validated across five studies by Donnellan et al. (2006) and showed consistent and acceptable internal consistencies across five studies ( $\alpha$  at or well above .60). The consistencies offer similar coverage as other Big Five measures as well as test-retest correlations similar to the parent measure across intervals of a few weeks and also several months (Donnellan et al., 2006). The Mini-IPIP scales produced acceptable reliability, especially for the reduced length. These coefficients ranged from .65 for Intellect/Imagination

to .93 for Extraversion. They also computed the associations between the Mini-IPIP scales and 6-items (from the 10-item “parent” IPIP-FFM scale) not included in the scale, and these correlations were also high, ranging from .78 for Extraversion to .56 for Intellect/Imagination. In addition, the Mini-IPIP scales showed a comparable pattern of convergent, discriminant and criterion-related validity with other Big Five measures. The predictive validity was also calculated and compared between the Mini-IPIP and other longer Big Five measures (BFI, IPIP-FFM) by regressing each of the three criteria on a given set of Big Five scales and recording the Multiple *R* values. In general the multiple *R* values were very similar across the three Big Five measures (compared to BAS, BFI  $R=.55$ , IPIP-FFM  $R=.60$ , Mini-IPIP  $R=.54$ ). Cooper, Smillie and Corr (2009) conducted a confirmatory factor analysis of the Mini-IPIP to demonstrate the suitability of the Mini-IPIP as a short-form measure of the FFM. Their results highlighted that the Mini-IPIP has acceptable reliability and a clearly interpretable factor structure (Cooper et al., 2009). The results from this study supported the original data which was provided by Donellan et al. (2006). In collaboration these results show that the Mini-IPIP is a psychometrically, acceptable and a practically useful short measure of the Big Five factors of personality. The Mini-IPIP has also been used in other recent studies (Grant & Wrzesniewski, 2010).

Participants read each of the items, which were formulated in sentence fragment form (e.g., “Am the life of the party”) and then rated how well it described them on a 5-point scale ranging from “very inaccurate” to very “accurate”. The Mini-IPIP measure used for this study can be viewed in Appendix A.

### **3.5.2 Task-orientated performance rating**

The dependent variable used in this study is the students' rating of the task-orientated performance of the lecturers. As defined earlier, task performance is "the effectiveness with which job incumbents' perform activities that contribute to an organisation's technical core either directly or by implementing a part of its technological process, or indirectly by providing it with needed materials or services" (Borman & Motowidlo, 1997, p. 99). Task activities that vary across different occupations are more likely to be role-prescribed and to involve cognitive ability. This was measured by the section focusing on the lecturers' performance on the annual evaluation forms used by the tertiary institution. The role-specific tasks expected to be done by lecturers is seen on the evaluation form.

Raters should be cooperative and trained in the techniques of rating, but they should also have direct experience with, or first-hand knowledge of, the individual to be rated (Cascio & Aguinis, 2005). It could be argued that the students who acted as raters in this study did have direct experience with the ratees, as they attended the lecturer's classes for at least one semester of a year. They did, however not have training in the technique of rating, which is a limitation to the current research. Rating scales typically have one of three types of anchors: numerical, adjectival or behavioural (Landy & Farr, 1980). Relative effectiveness of behavioural anchors was found in early studies, compared to numerical or adjectival anchors (Landy & Farr, 1980). The performance rating measure used for this research had a behavioural anchor as the items on the evaluation forms focused on behavioural aspects of the ratee.

The rating method has been shown to have an effect on the accuracy and utility of the performance rating information; whether one uses direct rating (in which the rater actually assigns a number to a ratee representing some level of performance) or methods of derived rating (in which the rater makes a series of discrete judgements about the ratee, from which a performance rating can be derived). The rating method chosen in this instance can be viewed as a derived rating as the rating sheets completed do not assign a number, but rather a discrete judgement from which a performance rating can be derived.

Participants read each of the items, which are formulated in sentence fragment form (e.g. “the lecturer was well prepared”) and then rated how well it described the lecturer on a 5-point scale ranging from “Disagree strongly” to “Agree strongly”. The items on the annual evaluation form used for the current study are listed in Appendix B.

### **3.5.3 Demographics**

Students (raters) and lecturers (ratees) completed basic demographic characteristic information in a last section of their respective questionnaires. The demographic information included age, gender, ethnic group, year of study, and course module. This information was collected so that the sample could be explicated in more detail. Furthermore, with the purpose of providing a strong test for the hypotheses this information was included in order to enable the researcher to control for possible extraneous influences in the study, should it be deemed necessary.

### **3.6 Procedure**

The selected classes and lecturers were informed of the research and the significance thereof was explained to them. Thereafter participants received an informed consent form to sign if they were willing to partake in the research. Each student (rater) completed a brief self-report personality measure and a demographic information sheet together with the semester module evaluation form. In order for responses to remain strictly confidential, the personality test, demographic information sheet and evaluation form of each student was coded by a number to link all the forms for a participant to one another. Each lecturer (ratee) received the same personality questionnaire for completion. Participation in the study was completely voluntary and emphasis was put on the confidentiality of the use of the captured data.

### **3.7 Data Analysis**

This section explains the method of analysis used to test the hypotheses formulated for this study, namely polynomial regression with response surface analysis. Due to its novelty, a more detailed explanation of this technique was provided in the literature review section.

Before the statistical analysis of the data, all questionnaires were checked for possible incorrect completion by respondents (i.e., the questionnaires were visually inspected and checked for unanswered questions or seemingly suspect answers). The data was analysed by applying various quantitative techniques. Firstly, descriptive statistics were presented and analysed, which, according to Gravetter and Forzano (2009), helps the researcher to organise, summarise, and simplify the results obtained from data collection. Specifically, descriptive

statistics in the form of means, standard deviations, skewness, kurtosis, and normality statistics of the study variables were computed.

Secondly, assumptions of form and distribution of data underlying the statistical techniques were considered. The psychometric measurement properties of the instruments used for the study were assessed with item analysis (for internal consistency reliability) and exploratory factor analysis (to identify common underlying factors) using SPSS. According to Field (2009), item analysis is performed with the objective of identifying and evaluating those items that are most related to their own construct, as well as other associated or similar constructs. The item analysis furthermore provided information regarding the sensitivity of the different items. This enables researchers to detect potentially poor items that should be deleted from the dataset. Next, the purpose of factor analysis is to determine what measures are measuring the same thing, and to what degree they are doing so (Tabachnick & Fidell, 2001). In other words, by performing factor analysis, one seeks to identify how many constructs an instrument measures.

When arriving at the testing of the research hypotheses, various inferential statistical techniques were used (in this case, correlation and polynomial regression with response surface analysis). Correlations were computed with the intention to assess the relationships between the dependent variables and the independent variables to see whether support was found for hypotheses 1 to 3. To test the similarity effects of the Big Five personality similarities (hypotheses 4 to 8), five polynomial regressions with response surface analysis were applied.



Regression analysis is one of the most commonly used techniques in statistics and is about exploring the association between dependent and independent variables (Fan & Gijbels, 1996). This is to assess the contribution of the independent variables and to identify the impact it has on the dependent variable. Linear regression is one of the most classical and widely used regression techniques. It will, however, create a very large modeling bias if the scatter plots examined for validation do not appear linear. A popular remedy for this problem is to increase the number of parameters by using polynomial regression, which was adopted for the present study.

### **3.7.1 Step-by-step overview of polynomial regression with response surface analysis**

A step-by-step overview of the polynomial regression with response surface analysis as presented by Shanock et al. (2010) is presented in the following section to shed more light on this technique (Shanock et al., 2010; Edwards, 1994, 2002).

With the response surface analysis approach, polynomial regression is conducted first. The general form of the equation to test for relationships using polynomial regression is  $Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e$ , where  $Z$  is a dependent variable,  $X$  is Predictor 1 (rater personality), and  $Y$  is Predictor 2 (ratee personality). The outcome variable is regressed on each of two predictor variables ( $X$  and  $Y$ ), the interaction between the two predictor variables ( $XY$ ), and the squared terms for each of the two predictors ( $X^2$  and  $Y^2$ ).

Instead of interpreting the direct results from the polynomial regression analysis, the coefficients from the analysis are used to examine what is called the “response surface

pattern” which is graphed to provide a three-dimensional visual representation of data to assist the interpretation.

The following steps are conducted when executing polynomial regression with response surface analysis:

*Step 1: Descriptive information about the occurrence of support discrepancies*

Before conducting the polynomial regression analyses, it is important to inspect how many participants would be considered to have discrepancies (differences) between the two predictors to get an idea of the base rate of discrepancies in the sample. Without this information, one would have no idea whether there were even discrepancies in your sample, how many and in what direction.

*Step 2: Run polynomial regression in SPSS and calculate the surface values*

Once it has been determined that discrepant values exist in the sample, the polynomial regression analysis can be conducted by firstly centering the predictors on the centered midpoint of their respective scales. After this, three new variables are made: (a) the square of the centered Predictor 1 variable; (b) the cross-product of the centered Predictor 1 and Predictor 2 variable; and (c) the square of the centered Predictor 2 variable. After this, the polynomial regression analysis is run.

*Step 3: Graphing the results in Excel*

In order to enhance the interpretation of the results, the three-dimensional response surface is plotted and its features examined. This is done by using the graphing functions in Excel.

*Step 4: Interpreting the surface values and graph*

Lastly, the graph and the calculated surface values are interpreted. Three things can be interpreted: (1) how the agreement in Predictor 1 and Predictor 2 relates to the outcome [PR]; (2) how the degree of discrepancy between Predictor 1 and Predictor 2 relates to the outcome [PR]; and (3) how the direction of Predictor 1 and Predictor 2 relates to the outcome [PR].

Polynomial regression analysis came about as a solution for the problems associated with using different scores to analyse discrepancies (differences) in ratings. There are several benefits of polynomial regression compared to difference scores (D). The first advantage of using polynomial regression instead of difference scores is that by combining distinct measures into one score, difference scores confound the effects of each of the component measures on the outcome.

### **3.8 Conclusion**

In chapter 3, the plan and methodology for the research process was explained. This includes a description of the applied research design, formulation of hypotheses, sample design and characteristics, information regarding the measuring instruments, and the way in which the data was collected. Finally, an outline of the different statistical techniques used to analyse the data was presented. The primary analysis techniques applied to test the formulated statistical hypotheses were correlation and polynomial regression with response surface analysis. Consequently, the results of the data analyses are presented in chapter 4.

## CHAPTER 4

### RESULTS

#### 4.1. Introduction

The researcher used a correlational design to test whether personality similarity between rater and ratee would have an effect on the rating of task-orientated performance ratings. Strong personality similarity in a rater-ratee relationship was hypothesised to have a positive effect on the performance rating conducted. Data was gathered in a laboratory study by using the performance rating that students conduct in respect of their lecturers at the tertiary institution used for the study. The data was analysed using polynomial regression with response surface analysis.

This chapter explains the way in which the data was cleaned after being captured in SPSS, whether there were missing values and how this was dealt with, and the way in which the assumptions for analysis were tested. Thereafter, the measurement properties of the dependent variable (performance rating) were assessed. As a main analysis, polynomial regression with response surface analysis was conducted and is reported.

#### 4.2. Data cleaning

The dataset was prepared for analysis by scanning and correcting for coding errors, missing values and outliers. This section explains the way in which the data was cleaned as deemed necessary by the researcher.

Accuracy of the data was ensured by cross-checking a random sample of ten per cent of the captured data with original questionnaires. There was, however, a few missing values identified when missing values analysis was conducted through SPSS 19.0. These missing values were deleted from the dataset as they represented less than 5% of the data. Frequency tables were requested and visually inspected. All values were found to be in range. Next, variable scores were standardised (z-scores) and inspected for possible extreme values (univariate outliers). No cases were significant ( $z < -3.29$  or  $z > 3.29$ ). Thus there were no univariate outliers. Lastly, there were certain negatively coded items on the Mini-IPIP questionnaire. These items were reflected and recoded for rater as well as ratee datasets.

### **4.3. Testing for assumptions**

In this section the general assumptions for polynomial regression are explained and reported. As mentioned in chapter 3, these assumptions are the same as for multiple regression analysis, namely normality, linearity and homoscedasticity.

#### **4.3.1. Normality, linearity, homoscedasticity**

Normality refers to the degree in which each variable and all linear combination of the variables are normally distributed and independent (Tabachnick & Fidell, 2007). Two of the components of normality are skewness and kurtosis, which can be tested for using statistical as well as graphical methods. Skewness and kurtosis were tested for by using SPSS descriptives and the data was found to be reasonably normal. Figures 4.1-4.5 depict the skewness and kurtosis visually, as histograms of the independent variables.

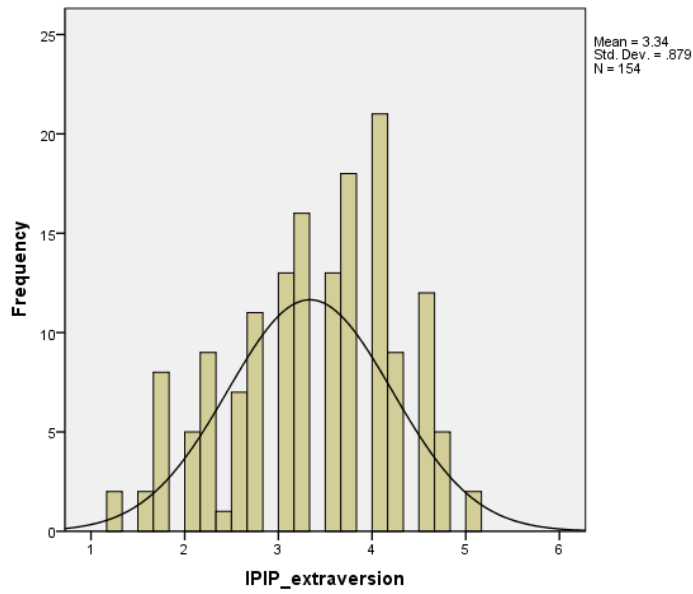


Figure 4.1. Histogram depicting normality of distribution for the IPIP EXTRAVERSION independent variable

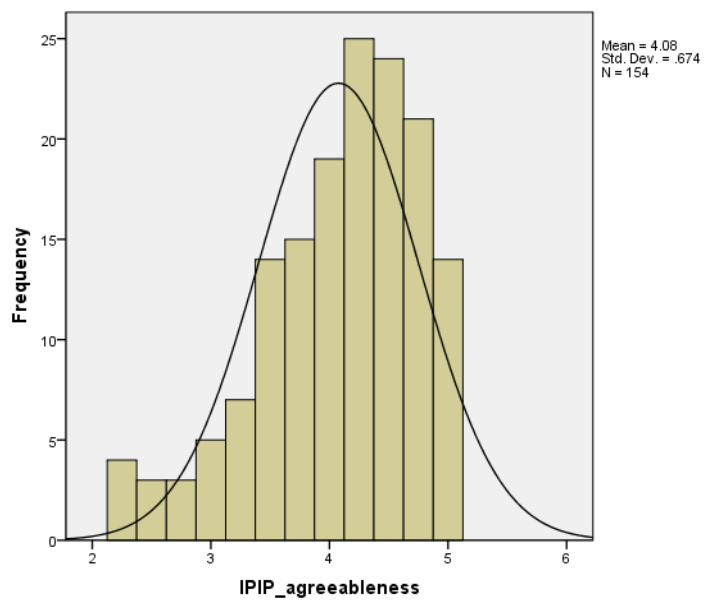


Figure 4.2. Histogram depicting normality of distribution for the IPIP AGREEABLENESS independent variable

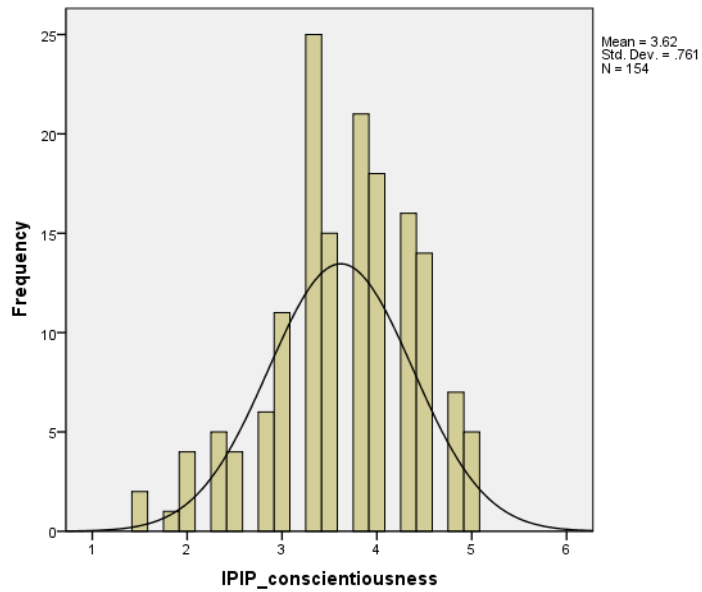


Figure 4.3. Histogram depicting normality of distribution for the IPIP CONSCIENTIOUSNESS independent variable

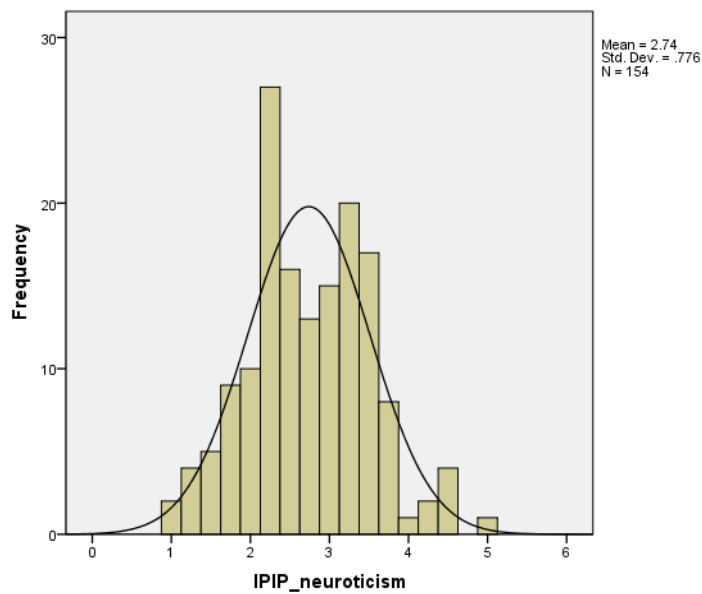
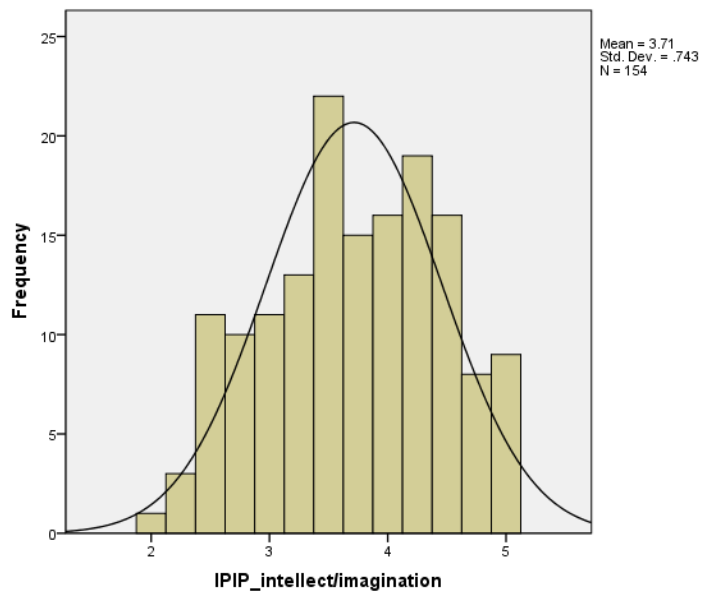


Figure 4.4. Histogram depicting normality of distribution for the IPIP NEUROTICISM independent variable



*Figure 4.5.* Histogram depicting normality of distribution for the IPIP OPENNESS independent variable

If there is a straight-line relationship between two variables (where one or both of the variables can be combinations of several variables), the linearity assumption is met (Tabachnick & Fidell, 2001). Linearity between two variables is assessed roughly through inspecting bivariate scatterplots through plotting the dependent and independent variables against each other.

Homoscedasticity relates to the assumption of normality, but considers whether the variability in scores for one continuous variable is roughly the same at all values of another continuous variable (Tabachnick & Fidell, 2001). A visual inspection was conducted on the scatterplots available in figures 4.6-4.8 and no obvious instances of non-linearity or heteroscedasticity were observed.



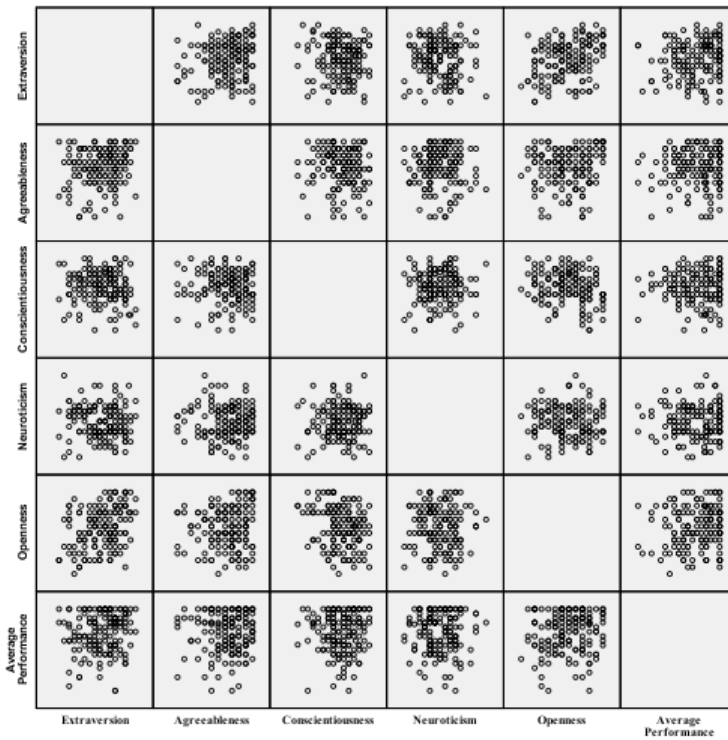


Figure 4.6. Graph output showing within-group scatterplot of rater personality dimensions and Average Performance.

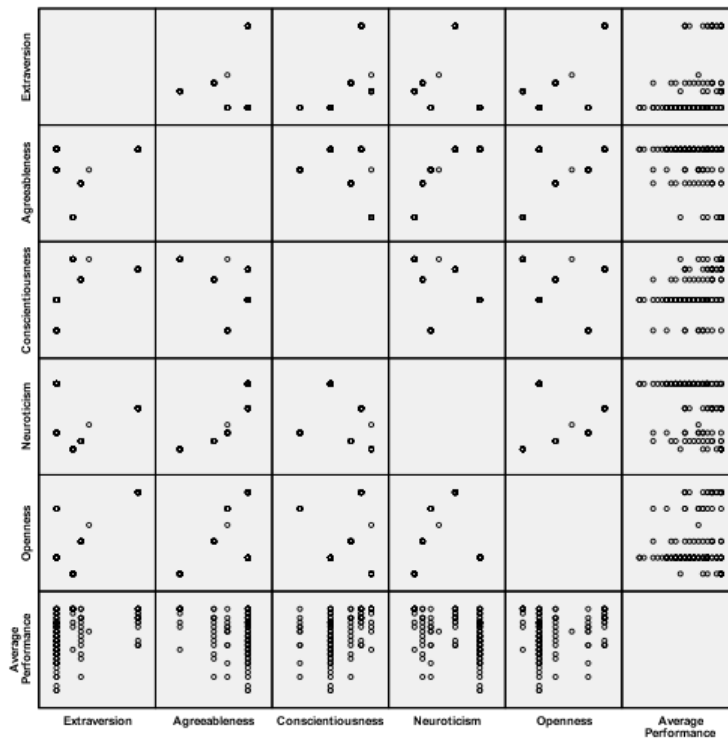


Figure 4.7. Graph output showing within-group scatterplot of rater personality dimensions and Average Performance.

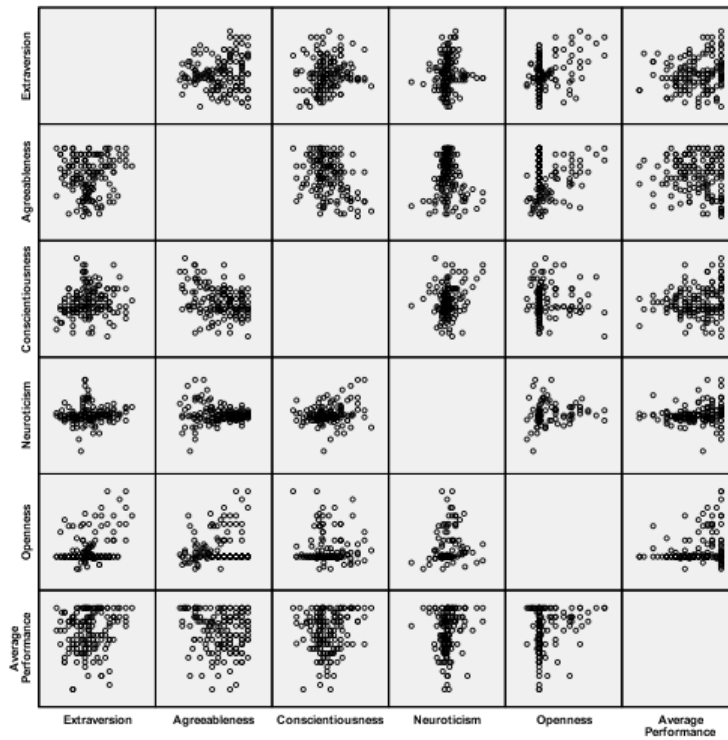


Figure 4.8. Graph output showing within-group scatterplot of rater-ratee interaction of personality dimensions and Average Performance.

#### 4.4. Assessing the measurement properties

A missing values analysis on performance rating items was conducted for all lecturers. No items had a substantial number of missing values (0.6 – 1.3%). Imputation was therefore not deemed necessary and this data was used for item analysis and exploratory factor analysis to further assess the measurement properties of the survey questionnaire.

#### **4.4.1. Subscale level descriptive**

##### **4.4.1.1. Item analysis**

An item analysis was conducted on all the items of the performance rating scale. The item analysis results were inspected and it was found that the Cronbach's alpha, also known as the internal consistency reliability estimate ( $\alpha$ ), was .882 on 11 items. This is considerably higher than the .80 cut-off for acceptable scale reliability of scores on a measure. When inspecting the items for scale summary item statistics, none of the items substantially affected the scale mean or scale variance. Based on these results of the item analysis for the performance rating scale, no items were identified for deletion; therefore, the same set of items could be factor-analysed with exploratory factor analysis (EFA).

##### **4.4.1.2. Exploratory factor analysis (EFA)**

The objective of the factor analysis is to explain the observed correlation matrix in terms of one or more common underlying factors. In other words, EFA analyses shared/common variance amongst the items (Tabachnick & Fidell, 2001). The intention was to analyse all the items as one factor, since a general performance score is needed. According to literature, there is one general factor underlying job performance (Viswesvaran, et al., 2005). Therefore, only one factor should be created for performance rating.

Dimensionality analysis (EFA) was conducted on all the performance measurement items and the results inspected. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .864 which is more than the .60 cut-off point, which means the correlation matrix was factor-

analysable. Next, the Bartlett's Test of sphericity, which showed a sig. value ( $p < 0.05$ ), was viewed, indicating that the null hypothesis could be rejected, indicating further support that the correlation matrix is factor-analysable.

The eigenvalues indicated that there were two factors present, which can be seen in table 4.1. The first factor explained 48.19% of the variance with an eigenvalue of 5.30, whilst the second factor only explained 12.02% of the variance with an eigenvalue of 1.32. The scree plot (figure 4.9), however, suggested that only one factor should be extracted as there was, conservatively speaking, one factor to the left of the inflection point on the plot. The eigenvalues and scree plot should be considered together for the decision of how many factors to extract. Although the Kaiser criterion indicates the existence of two factors, the scree plot suggests that one factor may be extracted. When inspecting the factor loadings the first broad factor represents the general performance factor and it was subsequently used to create the performance scores.

When inspecting the rotated component matrix, it was found that all items had factor loadings higher than .40, which was interpreted as the cut-off for significant factor loadings, except two items. These two items were PR\_5 and PR\_7, which had factor loadings on component one of .044 and .214 respectively (bold in Table 4.2). Both these items loaded very strongly ( $> .80$ ) on the second factor, which was interpreted as a feedback factor. Based on the results of the exploratory factor analysis, it was decided, firstly, to remove items PR\_5 and PR\_7, since they did not load on the general performance factor and, secondly, to retain the remaining items to create a subscale score on performance.

Reliability and exploratory factor analysis was conducted again after deletion of items PR\_5 and PR\_7. The internal consistency reliability estimate for the revised scale (without items PR\_5 and PR\_7) was .883 with nine items. The dimensionality analysis results showed that the eigenvalue for the general factor increased to 4.78, which explained 53.095% of variance. All items loaded on one factor with factor loadings ranging from .649 to .829.

Based on the item analysis and dimensionality analysis conducted on the performance rating items a total performance rating score was created for each rater by creating an unweighted linear composite of the performance rating items (PR\_1, PR\_2, PR\_3, PR\_4, PR\_6, PR\_8, PR\_9, PR\_10, PR\_11) with the SPSS TRANSFORM COMPUTE function. The averages of the item scores loading on the scale were used.

#### **4.4.2 Basic descriptive statistics of ratees**

Basic descriptive statistics of ratees can be viewed in Table 4.3. This is to indicate the difference in average ratings as well as distribution of scores with regards to ratings of each specific ratee.

After all of the above analyses, the separate datasets consisting of rater personality, ratee personality and performance ratings were merged so that each rater with his/her rating was linked to his/her appropriate ratee. Two cases [Ref numbers 56602 and 56872] were identified that could not be linked as they did not have performance ratings. These cases were deleted, bringing the number of dyad relationships to 152.

Table 4.1

*Total variance explained*

Component	Initial Eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	5.300	48.185	48.185	5.300	48.185	48.185	4.203	38.212	38.212
2	1.322	12.015	60.200	1.322	12.015	60.200	2.419	21.988	60.200
3	0.884	8.039	68.239						
4	0.698	6.346	74.585						
5	0.594	5.399	79.985						
6	0.526	4.782	84.766						
7	0.472	4.287	89.053						
8	0.433	3.934	92.987						
9	0.305	2.774	95.761						
10	0.275	2.503	98.263						
11	0.191	1.737	100						

*Note.* Extraction Method: Principal Component Method

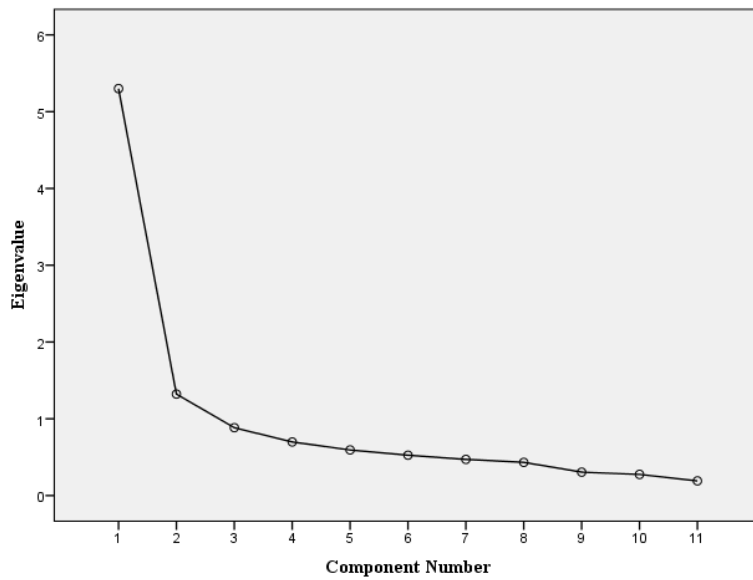


Figure 4.9 Scree-plot suggesting the amount of factors to extract from performance measure items.

Table 4.2

*Rotated component matrix of performance rating measure items*

	Component	
	1	2
Performance rating item		
PR_1	0.692	0.256
PR_2	0.472	0.562
PR_3	0.596	0.508
PR_4	0.797	0.256
PR_5	<b>0.044</b>	0.857
PR_6	0.698	0.119
PR_7	<b>0.214</b>	0.856
PR_8	0.751	-0.041
PR_9	0.778	0.244

PR_10	0.648	0.343
PR_11	0.622	0.229

Table 4.3  
*Descriptive statistics of each specific ratee's performance*

Lecturer/Ratee	Mean	N	Std. deviation
1	4.407	15	0.475
2	4.736	24	0.302
3	4.218	92	0.512
4	4.333	11	0.492
5	4.889	17	0.264
Total	4.394	154	0.519

#### 4.4.3 Means, standard deviations, and correlation between study variables

Means, standard deviations, alpha reliability ( $\alpha$ ), and inter-correlations for each of the study variables, dependent as well as independent, are represented in table 4.4. In general raters and ratees were similar in terms of the personality characteristics of extroversion, agreeableness and openness, but not in terms of conscientiousness and neuroticism as can be seen in table 4.4.

The relationship between rater personality traits and performance rating produced no significant correlation, which indicates that in this sample, in general, the personality of the rater does not influence the way in which they rate the performance of the lecturer. The correlation matrix, however, shows that there is a ratee main effect on performance rating. In



other words, lecturers who rated themselves high on, for example, extroversion was also assigned a high performance rating. It seems as if students preferred lecturers who are high in extroversion, conscientiousness and openness and low in agreeableness and neuroticism. It should, however, be taken into account that the number of ratees is small.

#### **4.5 Testing main effects**

Correlation was used with the intention to assess the main effects for hypotheses 1-3. Hypothesis one postulated that there is a statistically significant relationship between ratee extroversion and performance rating. As indicated in table 4.4 a positive and significant relationship ( $r = .357$ ;  $p$  (two tailed)  $< .01$ ), between the two variables was found. Furthermore, approximately 12.7% ( $r^2 = 0.127$ ) of the variance in the average performance rating can be explained in terms of the variance in extroversion.  $H_{01}$  can consequently be rejected in favour of  $H_{a1}$ , thus confirming hypothesis one.

Next, hypothesis two postulated that there is a statistically significant relationship between ratee conscientiousness and performance rating. As indicated in table 4.4, a positive and significant relationship ( $r = .413$ ;  $p$  (two tailed)  $< .01$ ) between the two variables was found. Furthermore, approximately 17.1% ( $r^2 = 0.170$ ) of the variance in the average performance rating can be explained in terms of the variance in conscientiousness.  $H_{02}$  can consequently be rejected in favour of  $H_{a2}$ , thus confirming hypothesis two.

Table 4.4

*Means, standard deviations, and correlations of study variables*

	M	SD	$\alpha$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
<u>Rater (Student)</u>														
(1) Extraversion	3.34	.884	.765											
(2) Agreeableness	4.07	.677	.735	.177*										
(3) Conscientiousness	3.62	.765	.646	-.069	-.075									
(4) Neuroticism	2.73	.778	.641	-.095	-.007	.043								
(5) Openness	3.71	.743	.698	.333**	.099	-.236**	-.027							
<u>Ratee (Lecturer)</u>														
(6) Extraversion	2.52	.896	-	.199*	-.036	.073	-.116	.139						
(7) Agreeableness	4.31	.812	-	.019	.186*	-.235**	-.091	.135	.008					
(8) Conscientiousness	4.22	.465	-	.048	-.122	.140	-.062	-.049	.661**	-.500**				
(9) Neuroticism	2.64	.775	-	-.081	.209**	-.236**	-.048	.031	-.309**	.880**	-.480**			
(10) Openness	3.42	.821	-	.242**	.007	.021	-.094	.216**	.796**	.262**	.101	-.199*		
(11) Average Performance	4.41	.486	.883	.116	.030	.023	.006	.101	.357**	-.307**	.413**	-.348**	.178*	

Note. \*  $p < 0.05$ , correlation is significant at the 0.05 level (2-tailed).

\*\*  $p < 0.01$ , correlation is significant at the 0.01 level (2-tailed).

$N = 152$

Lastly, hypothesis three postulated that there is a statistically significant relationship between rater openness and performance rating. As indicated in table 4.4, a positive and significant relationship ( $r = .178$ ;  $p$  (two tailed)  $< .05$ ) between the two variables was found. Furthermore, approximately 3.2% ( $r^2 = 0.031$ ) of the variance in the average performance rating can be explained in terms of the variance in openness.  $H_{03}$  can consequently be rejected in favour of  $H_{a3}$ , thus confirming hypothesis three.

#### 4.6 Frequency of Big 5 personality similarity and dissimilarity in sample (%)

Before polynomial regression was conducted the number of participants consisting of discrepancies between the two predictors was inspected. Without this information one would not know whether discrepancies even exist in a sample. It is noticeable, in table 4.5, that the amount of dissimilarity is quite high (61.8-75%) which means there is big variance in the dyadic relationships and it therefore becomes important to explore the effect of similarity on performance ratings.

Table 4.5

*Frequencies of dyadic similarities between rater and ratee personality dimensions*

	Percentage dissimilar	Percentage similar
<u>Dyadic similarity</u>		
Extraversion	69.7%	30.3%
Agreeableness	61.8%	38.2%
Conscientiousness	73.7%	26.3%
Neuroticism	71.7%	28.3%
Openness	75.0%	25.0%

Descriptive information about the occurrence of support discrepancies was found satisfactory in that there was enough dissimilarity in personality dimensions between dyads. Therefore, polynomial regression was proceeded with.

#### **4.7 Results of polynomial regression analysis**

Polynomial regression with response surface analysis was conducted separately for each one of the five personality traits. Table 4.6 reports these results for the separate polynomial regression analysis for rater extraversion (RE), ratee extraversion (EE), rater agreeableness (RA), ratee agreeableness (EA), rater conscientiousness (RC), ratee conscientiousness (EC), rater neuroticism (RN), ratee neuroticism (EN), rater openness (RO), ratee openness (EO) and performance rating (PR).

In general one can see that all five of the models were statistically significant with  $R^2$  ranging from .131 ( $p = .001$ ) for agreeableness to .265 ( $p = .000$ ) for conscientiousness. According to the surface test values ( $a_1$ -  $a_4$ ) there seems to be high similarity as well as high dissimilarity between the independent variables. Three out of five  $a_1$  values, which portrays similarity, are significant with the other two being very close to significant ( $p = .097$  and  $p = .060$ ). Also four out of the five  $a_3$  values, which portrays dissimilarity, are significant.

The first personality dimension is discussed in somewhat more detail in order to present a more thorough explanation where after the remaining four dimensions will be discussed briefly.

Table 4.6

*Results of Polynomial Regression Analysis on Big-5 personality dimensions*

	Extroversion	Agreeableness	Conscientiousnes:	Neuroticism	Openness
<u>Unstandardised Beta Coefficient</u>					
X	0.029	-0.070	-0.054	0.033	0.078
Y	0.277**	-0.326*	-0.700*	-0.408**	-0.682**
XY	0.062	0.096	0.104	0.035	-0.034
X <sup>2</sup>	0.026	-0.025	-0.080*	0.053	0.122
Y <sup>2</sup>	-0.221**	0.122	0.489**	-0.153	0.386**
R <sup>2</sup>	0.182†	0.131**	0.265†	0.141†	0.203†
Adjusted R <sup>2</sup>	0.154	0.101	0.240	0.112	0.176
a <sub>1</sub>	0.31†	-0.40	-0.75	-0.37**	-0.60†
a <sub>2</sub>	-0.13	0.19	0.51**	-0.06	0.47†
a <sub>3</sub>	-0.25**	0.26	0.65*	0.44**	0.76†
a <sub>4</sub>	-0.19	0.24	0.67†	-0.17	0.23

Note. X = Rater personality dimension [Predictor 1]

Y = Ratee personality dimension [Predictor 2]

a<sub>1</sub> = Slope along X=Y

a<sub>2</sub> = Curvature along X=Y; if significant the relationship between ratings that are in agreement and PR is nonlinear

a<sub>3</sub> = Slope along X= -Y

a<sub>4</sub> = Curvature along X= -Y; if significant shows influence of degree of discrepancy on PR

\**p* < .05; \*\**p* < .01; †*p* < .001

N=152

### 4.7.1 Extraversion

Rather than directly interpreting the results from the polynomial regression analysis, the coefficients from the analysis are used to examine the “response surface pattern” which is graphed to provide a three-dimensional visual representation of the surface values seen in figure 4.1-4.5. The slope and curvature of the two lines represents the response surface pattern. Firstly, response surface analysis allows one to examine the way in which similarity in a personality dimension relates to performance rating (PR). The black line on the floor of the response surface (e.g., figure 4.10) depicts  $X = Y$ , which for this personality dimension is  $RE = EE$ . This line is known as the “line of perfect agreement” or similarity. The slope of this line ( $a_1$ ), when considered in relation to the outcome (PR), represents the way in which the similarity between two predictor variables relates to the outcome. That is, the slope of this line shows the different levels of outcome variables for people whose levels of the two predictor variables are essentially the same across the continuum from low ratings on both predictors to high ratings on both predictors. Should  $a_1$  be positive, the outcome variable (PR) increases as the two predictors increase. Should  $a_1$  be negative, the outcome variable (PR) decreases as the two predictors increase. The curvature of this line ( $a_2$ ) represents the linearity of the line of perfect agreement. Should  $a_2$  be significant, the relationship between ratings that are in agreement and the outcome is nonlinear.

Secondly, the dashed line observed on the floor of the response surface (e.g. figure 4.10) is often called the “line of incongruence” or discrepancy and, for this personality dimension, depicts  $RE = - EE$ . Significant curvature ( $a_4$ ) along this line shows how the degree of discrepancy between RE and EE may influence the outcome variable. Significant negative curvature would mean that

the outcome variable (PR) suffers more as the levels of the two predictor variables diverge. The slope ( $a_3$ ) along the line of incongruence shows the extent to which the direction of the discrepancy matters and relates to height in the outcome variable (PR). In other words it shows whether the outcome is potentially affected more when the discrepancy is in one direction  $RE > EE$  (positive significant slope) or the other  $RE < EE$  (negative significant slope).

For extraversion the quadratic model accounted for a significant proportion of the rating variance ( $R^2 = .182$ ;  $F(5, 146) = 6.503$ ;  $p = .000$ ). The slope along the  $RE = EE$  line is statistically significant ( $a_1 = .13$ ,  $p = .000$ ) and positive, suggesting that there is a difference in ratings between dyads where the rater-ratee pair was high on extraversion and dyads where the pair was low on this dimension. Therefore, the null hypothesis ( $H_{04}$ ) that relational similarity in extraversion between rater and ratee does not affect performance rating is rejected. Stated otherwise, rater-ratee dyadic similarity in extraversion does seem to affect performance ratings; and also that PR increases as RE and EE increase. It seems as if a ratee main effect can be expected when a downward slope is seen along the surface. This suggests that ratings were higher when both rater and ratee were highly extroverted than when both rater and ratee were less extroverted. One can see the green patch on the response surface where the highest ratings occurred. This is when both rater and ratee were medium to highly extroverted. The lowest ratings occurred for almost all levels of raters when ratees were low on extraversion. For the curvature along  $RE = EE$  the value was not significant ( $a_2 = -.13$ ,  $p = .151$ ) and therefore the relationship is linear.

For extraversion the curvature along  $RE = -EE$  is negative and not significant ( $a_4 = -.19$ ,  $p = .100$ ) suggesting that PR is not significantly influenced as RE and EE diverge in similarity from

one another. This is seen through the flatter surface of the graph. Lastly the slope along  $RE = -EE$  is negative and significant ( $a_3 = -.25, p = .033$ ), which suggests that when  $RE < EE$ , the outcome (PR) is affected more than the other way around.

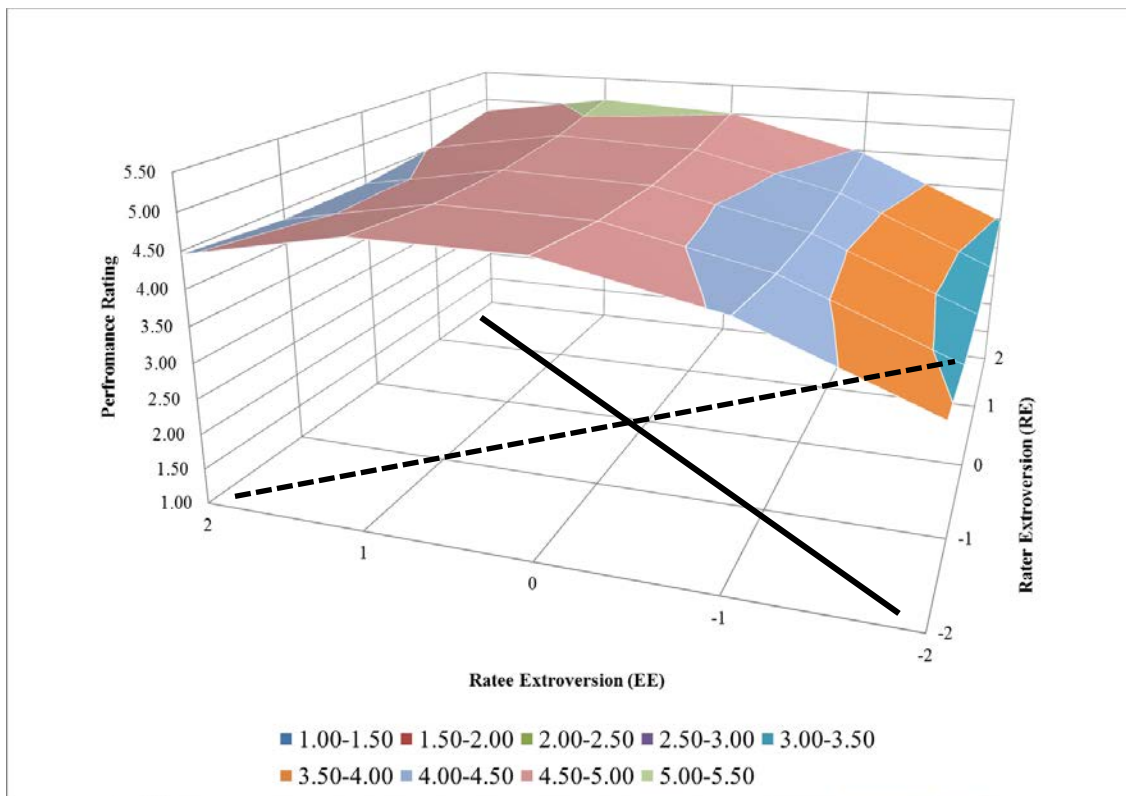


Figure 4.10 Performance rating as predicted by rater extraversion and ratee extraversion discrepancy.

### 4.7.3 Agreeableness

Figure 4.11 presents performance rating as predicted by rater agreeableness and ratee agreeableness discrepancy. The values of  $a_1$  to  $a_4$  can be seen in table 4.6. For the personality dimension of agreeableness the quadratic model accounted for a significant proportion of the



rating variance ( $R^2 = .131$ ;  $F(5, 146) = 4.406$ ;  $p = .001$ ). The slope along the  $RA = EA$  line is not significant, but very close ( $a_1 = -.40$ ;  $p = .097$ ) and negative, suggesting that the difference in ratings between dyads where the rater-ratee pair was high on extroversion and dyads where the pair was low on this dimension is almost significant. Therefore, the null hypothesis ( $H_{05}$ ) that relational similarity in agreeableness between rater and ratee does not affect performance rating cannot be rejected. It is however so close, that further inspection is done.

When inspecting the surface further, one sees that PR decreases as RA and EA increase. There is a tiny orange corner in the front right-hand side where the highest ratings occurred. This is where both rater and ratee are extremely low on agreeableness. The lowest ratings occurred when rater and ratee are high in agreeableness. For the curvature along  $RA = EA$  the value was not significant ( $a_2 = .19$ ,  $p = .125$ ) which means the relationship is linear. One can see the surface is only slightly concave. This is because of the very small  $a_2$  value.

With regard to the curvature along  $RA = -EA$ , it is positive and not significant ( $a_4 = .24$ ;  $p = .053$ ) suggesting that PR is not significantly influenced as RA and EA diverge in similarity from one another. This is seen through the flatter surface of the graph, but the slightly concave form is because of the positive value. Lastly the slope along  $RA = -EA$  is positive and not significant ( $a_3 = .26$ ,  $p = .101$ ), suggesting that the direction of discrepancy does not have a significant influence on PR.

### 4.7.3 Conscientiousness

Figure 4.12 presents performance rating as predicted by rater conscientiousness and ratee conscientiousness discrepancy. The values of  $a_1$  to  $a_4$  can be seen in Table 4.6. For the personality similarity of dimension conscientiousness the quadratic model accounted for a significant proportion of the rating variance ( $R^2 = .265$ ;  $F(5, 146) = 10.519$ ;  $p = .000$ ). The slope along the RC = EC line is negative and not significant ( $a_1 = -.75$ ,  $p = .060$ ). Therefore, the null hypothesis ( $H_{06}$ ) that relational similarity in conscientiousness between rater and ratee does not affect performance rating cannot be rejected. Even though the value is not significant, the fact that it is negative suggests that PR decreases as RC and EC increases. For the curvature along RC = EC the value was positive and significant ( $a_2 = .51$ ,  $p = .003$ ) and therefore the relationship is nonlinear.

For conscientiousness the curvature along RC = -EC is positive and significant ( $a_4 = .67$ ;  $p = .000$ ) suggesting that PR increases more sharply as the levels of RC and EC diverge. Lastly the slope along RC = -EC is also positive and significant ( $a_3 = .65$ ,  $p = .050$ ) which suggests that when RC > EC, the outcome (PR) is affected more than the other way around. One can see where the highest ratings occurred. The lowest ratings occurred where ratee was medium to high on conscientiousness for all the rater levels. When there is similarity between rater and ratee, there is definitely an effect on PR, but it seems to be the opposite of what would generally be expected, because where both rater and ratee is high on conscientiousness PR is low and where both rater and ratee is low on conscientiousness, PR is high.

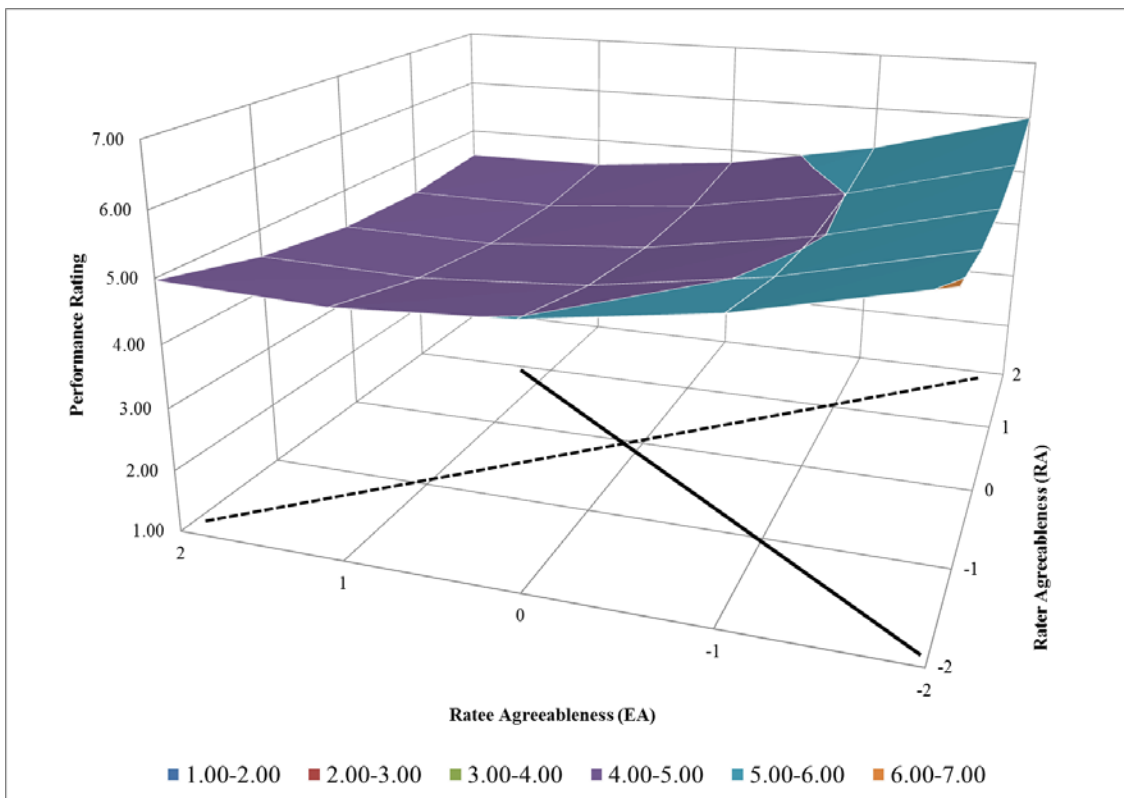


Figure 4.11 Performance rating as predicted by rater agreeableness and ratee agreeableness discrepancy.

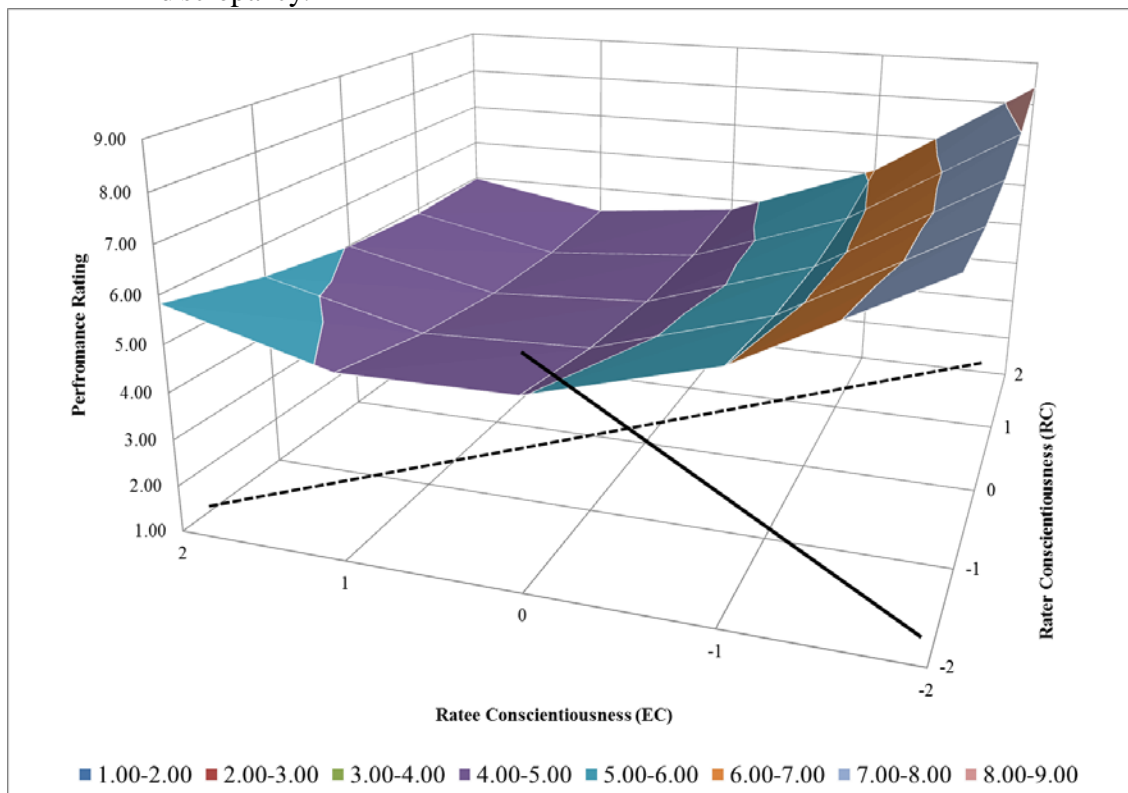


Figure 4.12 Performance rating as predicted by rater conscientiousness and ratee conscientiousness discrepancy.

#### 4.7.4 Neuroticism

Figure 4.13 presents performance rating as predicted by rater neuroticism and ratee neuroticism discrepancy. The values of  $a_1$  to  $a_4$  can be seen in table 4.6. For neuroticism the quadratic model accounted for a significant proportion of the rating variance ( $R^2 = .141$ ;  $F(5, 146) = 4.802$ ;  $p = .000$ ). The slope along the  $RN = EN$  line is statistically significant ( $a_1 = -.37$ ,  $p = .022$ ) and negative, suggesting that there is a difference in ratings between dyads where the rater-ratee pair was high on extroversion and dyads where the pair was low on this dimension. Therefore, the null hypothesis ( $H_{07}$ ) that relational similarity in neuroticism between rater and ratee does not affect performance rating, is rejected. Stated otherwise, rater-ratee dyadic similarity in neuroticism does seem to affect performance ratings.

With further inspection one can see that PR decreases as RN and EN increase. It seems as if a ratee main effect can be expected as a general upward slope is seen along the surface. This suggests that higher ratings were appointed by all raters, regardless of their position on neuroticism scale, when ratee's were low on neuroticism. One can see this in the pink section on the response surface. The lowest ratings occurred when ratees were high on neuroticism. For the curvature along  $RN = EN$  the value was not significant ( $a_2 = -.06$ ,  $p = .649$ ) and therefore the relationship is linear.

For neuroticism, the curvature along  $RN = -EN$  is negative and not significant ( $a_4 = -.17$ ,  $p = .230$ ), suggesting that PR is not significantly influenced as RN and EN diverge in similarity from one another. This is seen through the flatter surface of the graph. Lastly the slope along

$RN = -EN$  is positive and significant ( $a_3 = .44$ ,  $p = .007$ ) which suggest when  $RN > EN$ , the outcome (PR) is affected more than the other way around.

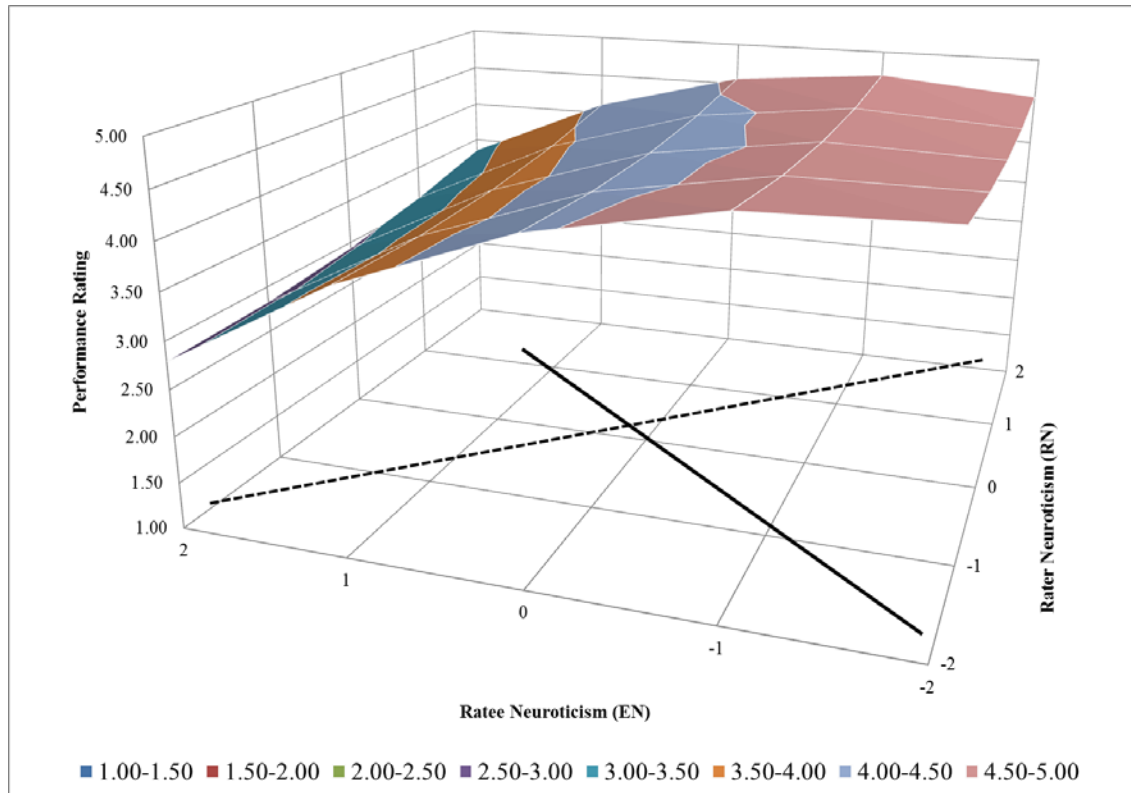


Figure 4.13 Performance rating as predicted by rater neuroticism and ratee neuroticism discrepancy

#### 4.7.5 Openness

Finally, figure 4.14 presents performance rating as predicted by rater openness and ratee openness discrepancy. The values of  $a_1$  to  $a_4$  can be seen in table 4.6. For openness the quadratic model accounted for a significant proportion of the rating variance ( $R^2 = .203$ ;  $F(5, 146) = 7.429$ ;  $p = .000$ ). The slope along the  $RO = EO$  line is negative and statistically significant ( $a_1 = -.60$ ,  $p = .000$ ), suggesting that there is a difference in ratings between dyads

where the rater-ratee pair was high on openness and dyads where the pair was low on this dimension. Therefore, the null hypothesis ( $H_{08}$ ) that relational similarity in openness between rater and ratee does not affect performance rating, is rejected. Stated otherwise, rater-ratee dyadic similarity in openness does seem to affect performance ratings.

Furthermore, one sees that PR decreases as RO and EO increase. It seems as if there is a ratee main effect as an upward slope is seen along the surface. This suggests that ratings were higher when both rater and ratee were low on openness than when both rater and ratee were more open. One can see the highest ratings on the right-hand side of the response surface compared to the lowest ratings on the left-hand side. For the curvature along  $RO = EO$  the value was significant ( $a_2 = .47, p = .000$ ) and therefore the relationship is nonlinear.

With regard to the curvature along  $RO = -EO$ , it is positive and not significant ( $a_4 = .23, p = .081$ ), suggesting that PR is not significantly influenced as RO and EO diverge in similarity from one another. Lastly, the slope along  $RO = -EO$  is positive and significant ( $a_3 = .76, p = .000$ ), which suggests that when  $RO > EO$ , the outcome (PR) is affected more than the other way around. This can also be seen clearly in the response surface, namely that where RC is high with a low EO (on the right-hand side of the graph), the PR is higher than where EC is high with a low RO (on the left-hand side of the graph).

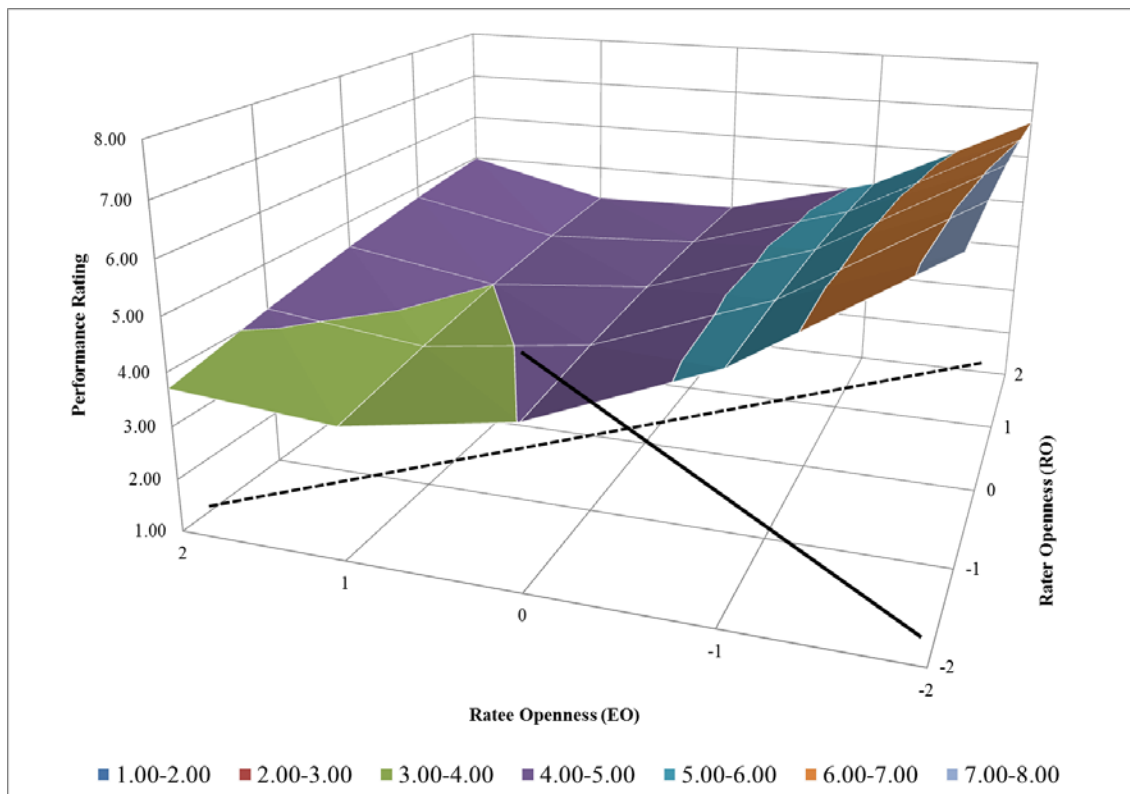


Figure 4.14 Performance rating as predicted by rater openness and ratee openness discrepancy.

#### 4.8 Summary and conclusion

The purpose of this chapter was, firstly, to prepare the data for analysis by testing the assumptions underlying multivariate procedures. The results from the analyses, i.e. correlation and polynomial regression with response surface analyses, which tested the proposed hypotheses developed in chapter 2, were then presented.

The results from the correlation analyses revealed significant relationships between ratee extraversion, conscientiousness and openness with average performance ratings. The strongest relationship was indicated between conscientiousness and average performance rating of the ratee ( $r = .413; p < .01$ ).

The results of the regression suggested that rater-ratee personality similarity with regards to extraversion, neuroticism and openness had a significant effect on the rating of performance. Even though rater-ratee personality in agreeableness did not have a significant effect on performance rating, it came very close to showing a significant effect ( $p = .097$ ). It is also very interesting to note that similarity in conscientiousness did not have an effect. The patterns response surfaces produced interesting information with regard to the effect that the direction of discrepancy has on the performance rating.

In summary, the results of these analyses were highly satisfactory, supporting six out of eight hypotheses. The next chapter discusses these findings in more detail and presents the limitations in the present study, whilst offering recommendations for future research on this topic.



## CHAPTER 5

### DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

#### 5.1 Introduction

This chapter presents a discussion of the general conclusions that were reached after the results of the data analysis had been interpreted. The results are summarised and interpreted within the context of previous research and theoretical frameworks. Next, the implications and conclusions drawn from the results as well as the limitations of the study are discussed. The chapter concludes with recommendations with regard to future research and possible practical applications of the results.

#### 5.2 Background of the study

Based on previous research relating to personality similarity in performance appraisal, the main concern of this study was to determine whether 1) differences exist in ratings assigned by raters (students) to ratees (lecturers) with specific personality traits (of the Big Five personality dimensions) with regard to task-orientated work behaviour, and 2) also to explore the extent in which rater-ratee personality similarity of any of the Big Five personality dimensions acts as one of the possible causes of systematic error in performance appraisal of task-orientated work behaviours.

The hypotheses of this study proposed that there are specific personality traits which are positively related to performance ratings, and also that similarity in personality between the

rater and ratee on the Big 5 personality dimensions is positively related to the performance rating of task-orientated work behaviours. If similarity effects are found, as hypothesised, it means that they most likely contribute to the error variance of the performance appraisal process and should be minimised to the extent that it is practically possible to do so. Having the knowledge of certain systematic interactions between rater and ratee characteristics will give practitioners the opportunity to develop better decision making systems.

To achieve the objectives set for the study, a correlational design was chosen with a sample consisting of final-year and postgraduate students who rated the performance of their lecturers. The evaluation forms were the formal feedback forms used by the tertiary institution on completion of each subject module. These evaluation forms focused on the behavioural actions of the lecturer that are more task-orientated than contextual orientated, as was the case with previous studies of this nature. Although the convenience sample selected represents a specific higher education environment, the focus should remain on the rating of task performance behaviour generically, and not on the sample's specific environment. The results were computed through polynomial regression with response surface analysis.

### **5.3 Summary of results**

As background to the discussion, this section briefly summarises the results obtained as well as other interesting findings, which are then be discussed. A summary of the stated hypotheses is provided in table 5.1. Significant positive effects for ratee extraversion, conscientiousness and openness on performance rating were found. For these main effects, openness was the lowest ( $r = .178$ ) and conscientiousness the highest ( $r = .413$ ).

For the five similarity hypotheses,  $H_{04}$ ,  $H_{07}$  and  $H_{08}$  were rejected, meaning that rater-ratee similarity in extraversion, neuroticism and openness had a significant effect on performance rating. With regard to agreeableness ( $H_{05}$ ) and conscientiousness ( $H_{06}$ ) the hypotheses could not be rejected. However, agreeableness came very close to being significant ( $r = .097$ ).

The results of these analyses were satisfactory in that they supported six out of the eight hypotheses as stated. These results are consequently integrated with prior literature and discussed.

#### **5.4 Discussion of results**

This section discusses the results by comparing them to prior literature generally, and the studies that were constructively replicated (Antonioni et al., 2001; Strauss et al., 2001) more specifically. Results are placed in the discussion section as facilitation to make it easier for the reader.

Results with regard to the main effects concurred with previous research, since significant positive effects for ratee extraversion, conscientiousness and openness on performance ratings were found. This supports the findings of a meta-analysis conducted by Tett, Jackson and Rothstein (1991), confirming that personality measures are predictors of job performance (extraversion:  $r = .13$ , conscientiousness:  $r = .155$ , openness:  $r = .236$ ), which was based on a review of 494 studies. In doing so, the outcome of this study is in support of other investigations as well (Barrick & Mount, 1991; Mount, Barrick & Stewart, 1998; Barrick, Stewart & Piotrowski, 2002).

In general, the similarity effects found in this study were more pronounced than that of Antonioni et al. (2001) as well as Strauss et al. (2001). Perhaps the fact that reliability affects validity contributed to this, since reliability in this study was higher than in theirs. In the present study it was shown that relational similarity in extraversion between rater and ratee dyad affects performance rating ( $R^2 = .182$ ). This contradicts the findings of Strauss et al. (2001) which found that relational personality for extraversion was *not* related to performance ratings by either supervisors ( $R^2 = .064$ ) or peers ( $R^2 = .065$ ). The fact that the lowest rating occurred where both rater and ratee was low on extraversion and the highest where both were high supports the theory of similarity attraction, which states that similar attitudes/personality/demographic characteristics are perceived as being rewarding, whereas dissimilar attitudes/personality/demographic characteristics are seen as negative reinforcements.

The finding of substantial personality similarity effects in this study contradicted the findings of earlier research (Antonioni et al., 2001) with regard to agreeableness, which found a significant relationship between similarity in agreeableness and three performance dimensions (consideration of others:  $R^2 = .041$ ; interpersonal communication:  $R^2 = .045$ ; and self-management responsibility:  $R^2 = .020$ ). The results of the present study could not support the hypothesis that a ratee's task performance will be higher when both rater and ratee have high levels of agreeableness than when both rater and ratee have low levels of agreeableness. Highly agreeable people may be more motivated to maintain positive relationships with other people than less agreeable people, as they are more tolerant, friendly and trustworthy. Individual differences in agreeableness are furthermore systematically associated with

Table 5.1

*Summary of hypotheses and conclusions*

Hypothesis	Reject/ Not reject	Conclusion
1. There is a statistically significant positive relationship between ratee extraversion and performance rating.	Reject H <sub>01</sub>	A significantly positive relationship exists between ratee extraversion and average performance rating.
2. There is a statistically significant positive relationship between ratee conscientiousness and performance rating.	Reject H <sub>02</sub>	A significantly positive relationship exists between ratee conscientiousness and average performance rating.
3. There is a statistically significant positive relationship between ratee openness and performance rating.	Reject H <sub>03</sub>	A significantly positive relationship exists between ratee openness and average performance rating.
4. Ratings of the ratee's task performance are higher when both rater and ratee have high levels of extroversion than when both rater and ratee have low levels of extroversion.	Reject H <sub>04</sub>	Relational similarity in extraversion between rater and ratee dyad affects performance rating.
5. Ratings of the ratee's task performance are higher when both rater and ratee have high levels of agreeableness than when both rater and ratee have low levels of agreeableness.	Not reject H <sub>05</sub>	Relational similarity in agreeableness between rater and ratee dyad does not affect performance rating.
6. Ratings of the ratee's task performance are higher when both rater and ratee have high levels of conscientiousness than when both rater and ratee have low levels of conscientiousness.	Not reject H <sub>06</sub>	Relational similarity in conscientiousness between rater and ratee dyad does not affect performance rating.
7. Ratings of the ratee's task performance are higher when both rater and ratee have low levels of neuroticism than when both rater and ratee have high levels of extroversion.	Reject H <sub>07</sub>	Relational similarity in neuroticism between rater and ratee dyad affects performance rating.
8. Ratings of the ratee's task performance are higher when both rater and ratee have high levels of openness than when both rater and ratee have low levels of openness.	Reject H <sub>08</sub>	Relational similarity in openness between rater and ratee dyad affects performance rating.

patterns of conflict and conflict resolution strategies. Perhaps the fact that ratings were conducted in a student-lecturer relationship had an influence on this outcome. As a student, one would not necessarily want a lecturer who agrees with everything you think, because a lecturer should challenge your views and ways of thinking. The results showed that similarity in agreeableness did, however, come very close ( $p = .097$ ) to showing a relationship with ratings.

Similarity in terms of conscientiousness does not always seem to affect performance ratings. For instance, in Antonioni et al. (2001) effects were found (consideration of others:  $R^2 = .027$ ; interpersonal communication:  $R^2 = .029$ ; and self-management responsibility:  $R^2 = .034$ ), whereas in the present study as well as in Strauss et al. (supervisors:  $R^2 = .055$ ; peers:  $R^2 = .094$ ) effects with regard to conscientiousness were not found. Furthermore, this relationship was found to be negative for the present study. In other words, performance rating decreased as rater and ratee conscientiousness increased. A high level of conscientiousness is associated with focus on goals and with being disciplined, responsible, and planful. It is also associated with a “will to achieve” (Kroeck & Brown, 2004). Low conscientiousness is related to procrastination, being unproductive, unorganised and irresponsible. Raters high in conscientiousness tend to expect similar behaviours from other people, will place more value on others’ conscientiousness, and will most likely develop more positive and constructive relationships. Perhaps higher-conscientious raters are more strict and objective in their rating, whereas less conscientious raters have a “never-minded” attitude toward rating. Nevertheless, similarity was found in the sense that raters who were low on conscientiousness rated similar ratees higher than those who were high.

Whereas conscientious similarity did not affect performance ratings, relational similarity in neuroticism between rater and ratee dyad affected performance rating in the present study, supporting the results of Strauss et al. (2001) for emotional stability ( $R^2 = .035$  for supervisor ratings of performance), which is depicted as neuroticism in the present study. Neuroticism is characterised by traits such as anxiety, depression, anger, embarrassment, and insecurity in contrast to emotional characteristics such as calmness, security and self-confidence. High neuroticism hinders effective interaction with others. The results make sense in that higher neurotics received lower ratings.

Relational similarity in openness between rater and ratee dyads also affected performance rating. The highest rating occurred when both rater and ratee were low on openness and the lowest rating was obtained when ratee was high on openness and rater was low on openness. This is interesting, but might also have to do with the job in this specific sample, since one would have more respect for a slightly rigid and strict lecturer.

The present study made several advances beyond the work of similar earlier research by addressing some of the shortcomings and previous recommendations, firstly, by making use of *actual* performance ratings which are used for evaluative purposes instead of performance ratings which were merely used for research purposes as in Antonioni et al. (2001) and Strauss et al. (2001). This gives more fidelity to this study. The results of the study show that even though raters are aware of the fact that their ratings will have an effect on the ratee because it is real ratings, personality similarity influences rating behaviour. Furthermore, the behaviours were assessed through a BOS of task performance, which showed satisfactory psychometric characteristics. The students were also used to these evaluations, because the sample consisted of final-year and postgraduate students who were, by this time, accustomed

to the evaluation form. Rating accuracy can also be assumed to be quite high as a good acquaintance exists between raters and ratees because students have interacted with lecturers for at least six months. The study makes use of upward rating, since the students are rating the lecturers, compared to downward rating or peer ratings used in the other studies. This is relevant, because sometimes subordinates are expected to rate their leaders or managers, for example, when 360° performance appraisal is conducted. In this study, task-performance behaviours were assessed compared to contextual performance behaviour. Specific jobs require specific attributes. It is speculated that this is a reason why many results are different from what was previously found. For example, in Antonioni et al. (2001) when measuring contextual behaviours for sales people, agreeableness was more important, whereas in this study concerning task-performance behaviours, conscientiousness is of more importance.

To conclude, the results of this investigation show that the similarity of personality between rater and ratee when rating performance is much more complex than previously expressed. It shows that similarity as well as dissimilarity plays a substantial role in the effect it has on the rating of performance.

## **5.5 Limitations of study**

A few limitations were identified for the study. Firstly, it should be taken into account that the number of ratees was very small, which could influence the results. Furthermore, the sample in total was not as large as originally expected and replications of the research should be performed on a number of additional samples in future research.



Secondly, the sample is very specific in that it consists of students rating lecturers. This is an upward rating, which may influence the perception of the ratees as they are rating someone more senior to themselves. This also limits the generalizability of the study into the workplace as the sample is in such a specific environment. Attempts were made to correct for limitations by selecting final-year and postgraduate students for the sample. This means that they at least have some experience in the ratings conducted, since this is done after every semester and year subject in all previous study years. Also, there was sufficient time to complete the rating.

Thirdly, some degree of common source bias (common method bias) exists (Podsakoff, MacKenzie, Lee & Podsakoff, 2003), because both rater personality and ratee performance were obtained from the same source. This was partially averted by obtaining ratings from more than one source, i.e., self-ratings for rater and ratee personality.

Lastly, performance ratings were negatively skewed, i.e., most lecturers had relatively high ratings. However, this might not be due to sampling error, because performance distribution amongst university lecturers in the sample organisation is relatively negatively skewed. For future studies, we recommend sampling performance rating targets with a more normal distribution.

## **5.6 Practical implications and recommendations for future research**

Despite the limitations, some recommendations can be made. Firstly, this study should be replicated, because it produced different results from previous studies conducted on this effect. Clearly, personality of both the rater and ratee plays an important role in task

performance ratings, i.e. main effects. This is nothing new (Barrick & Mount, 1991; Mount et al., 1998; Barrick et al., 2002). However, the present study showed that personality trait interaction could also influence performance rating. This type of effect can probably be described as construct-irrelevant, i.e. a form of bias, which should be avoided as far as possible.

Secondly, finding similarity effects leads one to question the source of these effects. Strauss et al. (2001), for instance, studied perceived similarity as well as the role of familiarity and liking. No support was found for familiarity as a moderator between relational and perceived similarity and moderate support for liking as a mediator of the perceived personality similarity-performance rating relationship. Therefore, familiarity does not seem to play a role. It has been suggested that the interaction between interpersonal similarity and performance ratings is complicated, which requires further exploration along with inclusion of other variables. Future research should seek to explore other potential causes for personality similarity effects on task-performance ratings.

With regard to recommendation for practice, raters should be instructed to focus on the specific traits they have to rate and informed also that possible other irrelevant factors (like personality similarity) may influence their ratings. Should performance ratings be required, multiple raters can be used to minimise this effect (Holzbach, 1987). In order to determine how this is all related to the evaluation of performance in a job setting, one should consider what task has to be accomplished. Formal as well as informal evaluations are widely used and can be considered (Feldman, 1981). The knowledge that rater-ratee personality similarity plays a role when conducting performance appraisal is the first step toward dealing with the problem.

## **5.7 Concluding remarks**

The important role that personality plays – especially in the form of Big Five – in work settings (Barrick & Mount, 1991; Johnson, 2003) has also been demonstrated in the present study. In more general research, the effect of rater personality on performance ratings has been demonstrated (Salgado, 1997; Tett, et al., 1991). In this study, it was shown that not only does the rater's personality seem to affect performance ratings, but also that the interaction between rater and ratee personality on most of the Big Five traits affected performance. This seems to support the theory of similarity-attraction, social identity theory and leader-member exchange theory.

From this research, important recommendations have been made with regard to practice as well as future research. It is believed that it was premature to discontinue research on the similarity effect and encourage studies to continue delving into this topic.

## LIST OF REFERENCES

- Aguinis, H., Henle, C., & Ostroff, C. (2001). Measurement in work and organizational psychology. In N. Anderson, D. Ones, H. Sinangil, & C. Viswesvaran (Eds. 1), *Handbook of industrial, work and organizational psychology*. (pp. 27-50). London: SAGE Publications.
- Antonioni, D. (1999). Predictors of upward appraisal ratings. *Journal of Managerial Issues*, *11*(1), 26-36.
- Antonioni, D., & Park, H. (2001). The effects of personality similarity on peer ratings of contextual work behaviors. *Personnel Psychology*, *54*, 331-361.
- Balzer, W. (1986). Biases in the recording of performance related information: The effects of initial impression and centrality of the appraisal task. *Organizational Behavior and Human Decision Processes*, *37*, 329-347.
- Barrick, M.R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1-26.
- Barrick, M.R., Stewart, G.L., Neubert, M.J., & Mount, M.K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*(3), 377-391.
- Barrick, M.R., Stewart, G.L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, *87*(1), 43-51.
- Bates, R. (2002). Liking and similarity as predictors of multi-source ratings. *Personnel Review*, *31*(5), 540-552.

- Bauer, T. N., Green, S. G., Bauer, T. N., & Bauer, T. N. (1996). Development of leader-member exchange: A longitudinal test. *The Academy of Management Journal*, 39(6), 1538-1567.
- Behn, R. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 99-110.
- Borman, W. (1997). 360 ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, 7(3), 299-315.
- Borman, W., & Motowidlo, S. (1997). Task Performance and Contextual Performance: The meaning for personnel selection research. *Human Performance*, 10(2), 99-109.
- Brown, R. (2000). Social Identity Theory: past achievements, current problems and future challenges. *European Journal of Social Psychology*, 30, 745-778.
- Byrne, D. (1971). *The Attraction Paradigm*. New York: Academic Press.
- Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, 22, 139-162.
- Cooper, A. J., Smillie, L. D., & Corr, P. J. (2010). A confirmatory factor analysis of the Mini-IPIP five-factor model personality scale. *Personality and Individual Differences*, 48(5), 688-691.
- Davis, T. & Landa, M. (1999). A contrary look at performance appraisal. *Canadian Manager*, Fall, 18-28.
- DeNisis, A., Cafferty, P., & Meglino, B. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance*, 33, 360-396.
- Dobbins, G. H., Cardy, R. L., & Truxillo, D. M. (1988). The effects of purpose of appraisal and individual differences in stereotypes of women on sex differences in performance ratings: A laboratory and field study. *Journal of Applied Psychology*, 73(3), 551-558.

- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment, 18*(2), 192-203.
- Edwards, J.R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Process, 58*, 51-100.
- Edwards, J.R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N.W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). San Francisco: Jossey-Bass.
- Engle, E. M., & Lord, R. G. (2011). Implicit theories, self-schemas, and leader-member exchange. *The Academy of Management Journal, 40*(4), 988-1010.
- Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman & Hall.
- Favero, J.L. & Ilgen, D.R. (1989). The effect of ratee prototypicality on rater observation and accuracy. *Journal of Applied Psychology, 19*(2), 932-946.
- Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology, 66*(2), 127-148.
- Feldman, J. (1994). On the synergy between theory and application: Social cognition and performance appraisal. In R. S. Wyer & T. K. Srull (Eds. 2), *Handbook of social cognition*. (pp. 339 – 398). New Jersey: Lawrence Erlbaum Associates.
- Field, A.P. (2009). *Discovering statistics using SPSS* (3<sup>rd</sup> ed.). Thousand Oaks: Sage Publications.
- Foti, R., & Lord, R. (1987). Prototypes and scripts: The effects of alternative methods of processing information on rating accuracy. *Organizational Behavior and Human Decision Processes, 39*(3), 318-340.

- Funder, D.C. (1991). Global traits: A neo-allportian approach to personality. *Psychological Science*, 2, 31-39.
- Goldberg, C. B. (2005). Relational demography and similarity-attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group & Organization Management*, 30(6), 597-624.
- Goldberg, L., Johnson, J., Eber, H., Hogan, R., Ashton, M., Cloninger, C. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84-96.
- Goldberg, C. B., Riordan, C., & Schaffer, B. S. (2010). Does social identity theory underlie relational demography? A test of the moderating effects of uncertainty reduction and status enhancement on similarity effects. *Human Relations*, 63(7), 903-926.
- Goodwin, C.J. (2003). *Research in psychology: Methods and design* (3<sup>rd</sup> ed.). United States of America: Hermitage Publishing Services.
- Grant, A. M., & Wrzesniewski, A. (2010). I won ' t let you down . . . or will I ? Core self-evaluations, other-orientation, anticipated guilt and gratitude and job performance. *Journal of Applied Psychology*, 95(1), 108 -121.
- Gravetter, F.J. & Forzano, L.B. (2009). *Research methods for the behavioral science*. (3<sup>rd</sup> ed.). USA: Wadsworth, Cengage Learning.
- Grobler, P.A., Wörnich, S. Carrell, M.R., Elbert, N.F. & Hatfield, R.D. (2011). *Human resource management in South Africa* (4<sup>th</sup> ed.). Singapore: Cengage Learning.
- Guion, R.M. (1998). *Assessment, measurement, and prediction for personnel decisions*. London: Lawrence Erlbaum Associates.
- Harris, C. (1988). A comparison of employee attitude toward two performance appraisal systems. *Public Personnel Management*, 17(4).
- Holzbach, R.L. (1978). Rater bias in performance ratings: superior, self- and peer ratings. *Journal of Applied Psychology*, 63(5), 579-588.

- Ilgén, D., Barnes-Farrell, J., & McKellin, D. (1993). Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes*, 54, 321-368.
- Jenkins, R. (2003). *Social identity*. London: Routledge.
- Johnson, J.W. (2003). Toward a better understanding of the relationship between personality and individual job performance. In M.R. Barrick & A.M. Ryan (Eds. 1), *Personality and work: Reconsidering the role of personality in organizations* (pp. 83-120). San Francisco: Jossey - Bass.
- Jordan, T. (2002). Performance Appraisals: Do they work?
- Kline, T., & Sulsky, L. (2009). Measurement and assessment issues in performance Appraisal. *Canadian Psychology*, 50(3), 161-171.
- Kraiger, K., & Ford, J. K. (1985). A meta-analysis of rater rater effects in performance ratings. *Journal of Applied Psychology*, 70(1), 56-65.
- Kroeck, G., & Brown, K.W. (2004). Work applications of the big five model of personality. In J.C. Thomas (Eds. 1), *Comprehensive handbook of psychological assessment: Industrial and organizational assessment* (pp.109 – 129). New Jersey: John Wiley & Sons, Inc.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87(1), 72-107.
- Landy, F.J., Barnes, J., & Murphy, K. (1987). Correlation of perceived fairness and accuracy of performance evaluation. *Journal of Applied Psychology*, 63, 751-754.
- Hoffman, B., Lance, C.E., Bynam, B., & Gentry, W.A. (2010). Rater source effects are alive and well after all. *Personnel Psychology*, 63, 119-151.
- Lebas, M. (1995). Performance measurement and performance management. *International Journal of Production Economics*, 41, 23-35.



- Levy, P., & Williams, J. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30(6).
- Lin, T., Dobbins, G., & Farh, J. (1992). A field study of race and age similarity effects on interview ratings in conventional and situational interviews. *Journal of Applied Psychology*, 77(3), 363-371.
- Macdonald, H. A., & Sulsky, L. M. (2009). Rating formats and rater training redux: A context-specific approach for enhancing the effectiveness of performance management. *Canadian Journal of Behavioural Science*, 41(4), 227-240.
- Motowidlo, S.J., & Van Scotter, J.R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475-480.
- Mount, M. K., Barrick, M. R., & Stewart, L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, 11(2/3), 145-165.
- Mount, M., Judge, A., Scullen, S., Sytsma, M., & Hezlett, S. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, (51), 557-577.
- Muchinsky, P. M., & Monahan, C. J. (1987). What is person-environment congruence? Supplementary versus complementary models of fit. *Journal of Vocational Behavior*, 27, 268-277.
- Murphy, K.R. & Cleveland, J.N. (1995). *Understanding performance appraisal: social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Noe, R.A., Hollenbeck, J.R., Gerhart, B., & Wright, P.M. (2006). *Human resource management: Gaining a competitive advantage* (5<sup>th</sup> ed.). United States: McGraw-Hill.
- Neely, A., Gregory, M., & Platts, K. (2005). Performance measurement system design: A literature review and research agenda. *International Journal of Operational, Production Management* 25(12), 1228-1263.

- Oppler, S. H., Campbell, J. P., Pulakos, E. D., & Borman, W. (1992). Three approaches to the investigation of subgroup bias in performance measurement: Review, results, and conclusions. *Journal of Applied Psychology*, 77(2), 201-217.
- Pervin, A.L., Cervone, D. & John, O.P. (2005). *Personality: Theory and research* (9<sup>th</sup> ed.). USA: John Wiley & Sons, Inc.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, 88(5), 879-903.
- Polk, J. (2006). Traits of effective teachers. *Arts Education Policy Review*, 107(4), 23-29.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2<sup>nd</sup> ed.). Thousand Oaks: Sage Publications.
- Republic of South Africa. (1998). *Employment Equity Act*. Government Gazette, 400 (19370), Cape Town.
- Roberts, G.E. (2003). Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, 30(1), 89-98.
- Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology*, 88 (5), 852-865.
- Sackett, P. R., & Dubois, C. L. (1991). Rater-ratee race effects on performance evaluation: Challenging meta-analytic conclusions. *Journal of Applied Psychology*, 76(6), 873-877.
- Salgado, J.F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82(1), 30-43.
- Schmitt, N., & Lippin, M. (1980). Race and sex as determinants of the mean and variance of performance ratings. *Journal of Applied Psychology*, 65(4), 428-435.

- Schmitt, N., Pulakos, E., Nason, E., & Whitney, D. (1996). Likability and similarity as potential sources of predictor-related criterion bias in validation research. *Organizational Behavior and Human Decision Processes*, 68(3), 272-286.
- Scullen, S., Mount, M., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, 85(6), 956-970.
- Shannock, L.R., Baran, B.E., Gentry, W.A., Pattison, S.C., & Heggestad, E.D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business Psychology*, 25, 543-554.
- Stevens, S. (1968). Measurement, statistics, and the schemapiric view. *Science*, 161(3844), 849-856.
- Strauss, J. P., Barrick, M. R., & Connerley, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology*, 637-657.
- Sundvik, L., & Lindeman, M. (1998). Performance rating accuracy: Convergence between supervisor assessment and sales productivity. *International Journal of Selection and Assessment*, 6(1), 9-15.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4<sup>th</sup> ed.). Needham Heights, MA: Allyn & Bacon.
- Tajfel, H. & Turner, J.C. (1986). The Social Identity Theory of intergroup behavior. In S. Worchel & W. Austin (eds.), *Psychology of Intergroup Relations* (pp. 7-24). Chicago: Nelson Hall.
- Taylor, S. & Fiske, S. (1991). *Social Cognition* (2<sup>nd</sup> ed.). England: McGraw-Hill Book Company.

- Taylor, N. (2004). *The construction of a South African five-factor personality questionnaire by Nicola Taylor*. Unpublished master's thesis, Rand Afrikaans University, Johannesburg.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance : A meta-analytic review. *Personnel Psychology*, *44*, 703-742.
- Turban, D. B., & Jones, a. P. (1988). Supervisor-subordinate similarity: types, effects, and mechanisms. *The Journal of Applied Psychology*, *73*(2), 228-34.
- Tziner, A., Murphy, K., Cleveland, J., Yavo, A., & Hayoon, E. (2008). A new old question: Do contextual factors relate to rating behavior: An investigation with peer evaluations. *International Journal of Selection and Assessment*, *16*(1).
- Varma, A., DeNisi, A.S., & Peters, L.H., (1996). Interpersonal affect and performance appraisal: A field study. *Personnel Psychology*, *49*(2), 341-360.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*, 557-574.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (2005). Is there a general factor in ratings of performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, *90*(1), 108-131.
- von Elverfeldt, A. (2005). *Performance appraisal – how to improve its effectiveness*. Retrieved from University of Twente Student Theses.
- Waldman, D. A., & Avolio, B. J. (1991). Race effects in performance evaluations: Controlling for ability, education, and experience. *Journal of Applied Psychology*, *76*(6), 897-901.
- Wayne, S.J., & Liden, R.C. (1995). Effects of impression management on performance ratings: A longitudinal study. *Academy of Management Journal*, *38*(1), 232-260.

Welman, C., Kruger, F., & Mitchell, B. (2005). *Research Methodology* (3<sup>rd</sup> ed.). Cape Town: Oxford.

Wexley, K., Alexander, R., Greenawalt, J., & Couch, M. (1980). Attitudinal congruence and similarity as related to interpersonal evaluations in manager-subordinate dyads. *Academy of Management Journal*, 23(2), 320-330.

Wherry, R., & Bartlett, C. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology*, 35, 521-552.

Winkler, I. (2009). *Contemporary leadership theories: Enhancing the understanding of the complexity, subjectivity and dynamic of leadership*. London: Physica-Verlag.

**APPENDIX A : Mini IPIP Questionnaire****MINI-IPIP QUESTIONNAIRE AND RATING SHEET****Instructions**

How Accurately Can You Describe Yourself?

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence.

Indicate for each statement whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you by circling the appropriate number.

Item	Very inaccurate	Moderately inaccurate	Neither accurate nor inaccurate	Moderately accurate	Very accurate
1. Am the life of the party.	1	2	3	4	5
2. Sympathize with others' feelings	1	2	3	4	5
3. Get chores done right away.	1	2	3	4	5
4. Have frequent mood swings.	1	2	3	4	5
5. Have a vivid imagination.	1	2	3	4	5
6. Don't talk a lot.	1	2	3	4	5
7. Am not interested in other people's problems.	1	2	3	4	5
8. Often forget to put things back in their proper place.	1	2	3	4	5
9. Am relaxed most of the time.	1	2	3	4	5
10. Am not interested in abstract ideas.	1	2	3	4	5
11. Talk to a lot of different people at parties.	1	2	3	4	5
12. Feel others' emotions.	1	2	3	4	5
13. Like order.	1	2	3	4	5
14. Get upset easily.	1	2	3	4	5
15. Have difficulty understanding abstract ideas.	1	2	3	4	5
16. Keep in the background.	1	2	3	4	5
17. Am not really interested in others.	1	2	3	4	5
18. Make a mess of things.	1	2	3	4	5
19. Seldom feel blue.	1	2	3	4	5
20. Do not have a good imagination.	1	2	3	4	5

THANK YOU

**APPENDIX B: Lecturer evaluation form**

	Verst sterk Disagree strongly	Verst Disagree	Neutral Neutral	Stem saam Agree	Stem volkom saam Agree Strongly	N.V.T. N/A
ix. Die fisiese onderrig- en leeruimtes (bv. klaskamer groottes, beligting, ens.) wat vir hierdie module gebruik is, was voldoende. / <i>The teaching and learning spaces (e.g. classroom size, lighting, etc.) used for this module was adequate.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
x. Die verwagte werkslading in hierdie module was billik. / <i>The amount of work required in this module was reasonable.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
xi. Die module was goed gestruktureerd. / <i>This module was well structured.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
xii. Hierdie module het my vaardighede in analise en oplossing van probleme help ontwikkel. / <i>This module has helped me develop my skills in analysis and problem solving.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>2. Addisionele kommentaar / Additional commentary</b>						
i. Wat was die beste aspekte van hierdie module? / <i>What were the best aspects of this module?</i>						
ii. Watter aspekte van hierdie module kan verbeter? / <i>What aspects of this module need improving?</i>						

<b>DOSENT / LECTURER :</b>						
	Verst sterk Disagree strongly	Verst Disagree	Neutral Neutral	Stem saam Agree	Stem volkom saam Agree Strongly	N.V.T. N/A
<b>1. Beoordeel die dosent se onderrig op 'n skaal van 1 tot 5: / Assess the lecturer's teaching on a scale of 1 to 5:</b>						
i. Die dosent was entoesiasies. / <i>The lecturer was enthusiastic.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ii. Die dosent se verduidelikings was duidelik. / <i>The lecturer's explanations were clear.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
iii. Die lesings was goed gestruktureerd. / <i>Lectures were well structured.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
iv. Die dosent was goed voorberei. / <i>The lecturer was well prepared.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
v. Terugvoer op take en toetsa het stipteik plaasgevind. / <i>Feedback on tasks and tests was given promptly.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
vi. Studente is aangemoedig om deel te neem aan module (bv. vrae te vra, voorstelle te maak). / <i>Students were encouraged to participate during the module (e.g. to ask questions, make suggestions).</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
vii. Studente het betekenisvolle terugvoer ontvang. / <i>Students received meaningful feedback.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
viii. Die dosent was toeganklik vir studente. / <i>The lecturer was accessible to students.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ix. Die dosent was intellektueel stimulerend. / <i>The lecturer was intellectually stimulating.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
x. Die dosent het my belangstelling in die onderwerp laat verdiep. / <i>The lecturer has increased my interest in the subject.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
xi. Studente in hierdie module is regverdig en met respek behandel. / <i>In this module, students were treated fairly and with respect.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>2. Addisionele kommentaar / Additional commentary</b>						
i. Aspekte van dosent se onderrigstyl wat behou moet word. / <i>Aspects of the lecturer's teaching style that should be maintained.</i>						
ii. Aspekte van dosent se onderrigstyl wat verbeter kan word. / <i>Aspects of the lecturer's teaching style that need improvement.</i>						

GEHEEL NDRUK: Aan die hand van jou terugvoer op die voorafgaande items, gee asseblief jou gemeelindruk van die module en ou deurdagte geheelindruk van die kwaliteit van die dosent se onderrig deur 'n persentasie toe te kan.  
 GENERAL IMPRESSION: Based on your feedback on the previous items, give your general impression of the module and your considered general impression of the quality of the lecturer's teaching expressed as a percentage.

Persentasie Module / Percentage Module    %      Persentasie Dosent / Percentage Lecturer    %

## **APPENDIX C: Consent form for ratee participation in research**

### **TERTIARY INSTITUTION CONSENT TO PARTICIPATE IN RESEARCH**

---

*Research title: The effect of rater-ratee personality similarity on the rating of task orientated work behaviours*

The proposed study is part of the Masters research of Ms. Antonette Crouse (Stellenbosch University, Department of Industrial Psychology) and plans to explore how raters go about to make judgements of others during performance appraisal. There are a number of factors that we are aware of that influence the ratings assigned to ratees. For this research project, you will be asked to complete a Mini-IPIP personality test.

#### **1. PURPOSE OF THE STUDY**

The purpose of the research study is to investigate the occurrence of personality similarity effects in the rating of performance.

#### **2. PROCEDURES**

If you volunteer to participate in this study, you will be required to complete the Mini-IPIP during the same time that the course-module evaluation form is completed by the students. The consent form will be filled in and collected where after the Mini-IPIP will be handed to you. We would ask you to do the following things:

##### **2.1. RECEIPT OF MINI-IPIP**

The Mini-IPIP will be handed out to the participants, in person, by the researcher.

##### **2.2. COMPLETION OF MINI-IPIP**

The participant is requested to read thoroughly and complete the Mini-IPIP individually as honest as possible. There are no right or wrong answers.

##### **2.3. QUESTIONNAIRE COLLECTION**

After completion, the researcher will personally collect the Mini-IPIP from participants immediately.

#### **3. POTENTIAL RISKS AND DISCOMFORTS**

There are no potential risks or discomforts envisaged in this study



#### **4. POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY**

Participants will gain knowledge on the performance appraisal process and the importance of an effective, valid and reliable process. Participants will also be given the opportunity to request personal feedback of the Mini-IPIP completed. Should they request this; an email can be sent to the researcher with the participants' reference number in order to draw the correct results for the feedback.

##### **4.1. BENEFIT TO SCIENCE/SOCIETY**

The need for the proposed research was initiated by the overarching need to understand the sources of variance in rater bias. Earlier research on similarity effects has mainly focused on demographical variables. But research has suggested that the basis of our similarity judgments changes in time. Research on the connection between rater-ratee personality similarity and ratings are quite sparse and has thus been identified as a fruitful area for performance ratings research.

#### **5. PAYMENT FOR PARTICIPATION**

No payment will be made to participants for partaking in this study.

#### **6. CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified will remain confidential and will be disclosed only with the participant's permission or as required by law. Confidentiality will be maintained by means of using coding procedures. The participants are not required to write their names or particulars on the questionnaires. The questionnaires will be issued to the participants by the researcher, in person. On completion of the questionnaire, the researcher will personally collect the questionnaire from the participant. The results will be published in the form of a completed dissertation as well as in an accredited journal, but confidentiality will be maintained. No names will be published.

#### **7. PARTICIPATION AND WITHDRAWAL**

You may choose whether or not to participate in this study. If you, the participant volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and still remain in the study. The researcher may withdraw you from this study if circumstances arise which warrant doing so, such as ill-health or resignation.

#### **8. IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact Antonette Crouse ([14834480@sun.ac.za](mailto:14834480@sun.ac.za) / 079 854 2558) Or Mr F. S. De Kock ([fsdk@sun.ac.za](mailto:fsdk@sun.ac.za) / 021 808 3016 / 082 780 4652)

#### **9. RIGHTS OF RESEARCH SUBJECTS**

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research subject, contact Ms Marlene Fouche ([mfouche@sun.ac.za](mailto:mfouche@sun.ac.za) / 021 808 4622) at the Division for Research Development, Stellenbosch University.

---

**Please tick the appropriate box below:**

A. I herewith give consent to participate in the research explained to me and as outlined above. I also agree that the data may be used for future similar research projects, without revealing my identity.

B. I do not give consent to participate in this research.

## **APPENDIX D: Consent Form for Rater Participation in Research**

### **TERTIARY INSTITUTION CONSENT TO PARTICIPATE IN RESEARCH**

---

***Research title: The effect of rater-ratee personality similarity on the rating of task orientated work behaviours***

The proposed study is part of the Masters research of Ms. Antonette Crouse (Stellenbosch University, Department of Industrial Psychology) and plans to explore how raters go about to make judgements of others during performance appraisal. There are a number of factors that we are aware of that influence the ratings assigned to ratees. For this research project, you will be asked to complete the standard course-module and lecturer evaluation form which is completed for every subject. In addition to that, you will be asked to complete a Mini-IPIP personality test.

#### **1. PURPOSE OF THE STUDY**

The purpose of the research study is to investigate the occurrence of personality similarity effects in the rating of performance.

#### **10. 2. PROCEDURES**

If you volunteer to participate in this study, you will be required to sit in class room for the duration of the data collection. The consent forms will be filled in and collected where after the standard course-module and lecturer evaluation form together with the Mini-IPIP will be distributed to you. We would ask you to do the following things:

##### **2.1. RECEIPT OF EVALUATION FORM AND MINI-IPIP**

The evaluation form and Mini-IPIP will be handed out to the participants, in person, by the researcher.

##### **2.2. COMPLETION OF EVALUATION FORM AND MINI-IPIP**

The participants are requested to read thoroughly and complete the evaluation form and Mini-IPIP individually and as honest as possible. There are no right or wrong answers.

##### **2.3. QUESTIONNAIRE COLLECTION**

After completion, the researcher will personally collect all evaluation forms and Mini-IPIP from participants immediately.

#### **11. 3. POTENTIAL RISKS AND DISCOMFORTS**

There are no potential risks or discomforts envisaged in this study

## **12. 4. POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY**

Participants will gain knowledge on the performance appraisal process and the importance of an effective, valid and reliable process. Participants will also be given the opportunity to request personal feedback of the Mini-IPIP completed. Should they request this; an email can be sent to the researcher with the participants' reference number in order to draw the correct results for the feedback.

### **4.1. BENEFIT TO SCIENCE/SOCIETY**

The need for the proposed research was initiated by the overarching need to understand the sources of variance in rater bias. Earlier research on similarity effects has mainly focused on demographical variables. But research has suggested that the basis of our similarity judgments changes in time. Research on the connection between rater-ratee personality similarity and ratings are quite sparse and has thus been identified as a fruitful area for performance ratings research.

## **6. 5. PAYMENT FOR PARTICIPATION**

No payment will be made to participants for partaking in this study.

## **6. CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified will remain confidential and will be disclosed only with the participant's permission or as required by law. Confidentiality will be maintained by means of using coding procedures. The participants are not required to write their names or particulars on the questionnaires. The questionnaires will be issued to the participants by the researcher, in person. On completion of the questionnaire, the researcher will personally collect the questionnaires from all participants. The results will be published in the form of a completed dissertation as well as in an accredited journal, but confidentiality will be maintained. No names will be published.

## **7. PARTICIPATION AND WITHDRAWAL**

You may choose whether or not to participate in this study. If you, the participant volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and still remain in the study. The researcher may withdraw you from this study if circumstances arise which warrant doing so, such as ill-health or resignation.

## **8. IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact Antonette Crouse ([14834480@sun.ac.za](mailto:14834480@sun.ac.za) / 079 854 2558) Or Mr F. S. De Kock ([fsdk@sun.ac.za](mailto:fsdk@sun.ac.za) / 021 808 3016 / 082 780 4652)

## 9. RIGHTS OF RESEARCH SUBJECTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research subject, contact Ms Marlene Fouche ([mfouche@sun.ac.za](mailto:mfouche@sun.ac.za) / 021 808 4622) at the Division for Research Development, Stellenbosch University.

---

**Please tick the appropriate box below:**

C. I herewith give consent to participate in the research explained to me and as outlined above. I also agree that the data may be used for future similar research projects, without revealing my identity.

D. I do not give consent to participate in this research.

## APPENDIX E: Ethical permission to conduct research



UNIVERSITEIT-STELLENBOSCH-UNIVERSITY  
jou kennisvernoot • your knowledge partner

### ETHICS REVIEW REPORT

---

Applicant:	Ms A Crouse
Project title:	The effect of rater-ratee personality similarity of task orientated work behaviours
Nature of research project:	MComm (Psych)
Supervisor (if applicable):	Mr F De Kock
Date:	24 September 2011

---

The research proposal of Ms A Crouse was considered and evaluated in terms of the guidelines prescribed by the Stellenbosch University Framework Policy to Promote and Ensure Ethically Responsible Research, adopted by the Senate on 20 March 2009. The research proposal was presented by the researcher during a formal presentation session on 23 September 2011 attended by Proff DJ Malan, A Engelbrecht, CC Theron, Drr G Gørgens and B Boonzaier and Ms S Adams . The purpose of this review is to ascertain whether there are any ethical risks associated with the proposed research project of which the researcher has to be aware or, to assess the nature and extent of these ethical risks, and to suggest measures that can be taken to avoid or minimize these risks.

#### Summary of Research

The objective of this study is [Application for ethical clearance, p. 1] "to investigate how the construct irrelevant variable, personality similarity between rater and rate, affects the ratings of task orientated work behaviours." The purpose of the research is to attempt to identify mechanisms that could be used to improve the validity of performance ratings.

#### Documents Received:

The Departmental Research Ethics Committee received the following documentation as part of the submission for ethical clearance:

An application for ethical clearance [Signed by the researcher, head of department and/or supervisor]	Yes
Copies of relevant letters of permission submitted	No <sup>1</sup>
Research Proposal	Yes

---

<sup>1</sup> A copy of an e-mail from Mrs E Pietersen from SCL was attached indicating SCL's willingness to provide performance ratings under the proviso that formal institutional permission should be obtained and that the proposed research be ethically cleared.